

Exact Dimensionality Reduction for Partial Line Spectra Estimation Problems

Maxime Ferreira Da Costa and Wei Dai

Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom

Email: {maxime.ferreira, wei.dai1}@imperial.ac.uk

Abstract

Line spectral estimation theory aims to estimate the off-the-grid spectral components of a time signal with optimal precision. Recent results have shown that it is possible to recover signals having sparse line spectra from few temporal observations via the use of convex programming. However, the computational cost of such approaches remains the major flaw to their application to practical systems. This work investigates the recovery of spectrally sparse signal from low-dimensional partial measurements. It is shown in the first part of this paper that, under a light assumption on the sub-sampling matrix, the partial line spectral estimation problems can be relaxed into a low-dimensional semidefinite program. The proof technique relies on a novel extension of the Gram parametrization to subspaces of trigonometric polynomials.

The second part of this work focuses on the analysis of two particular sub-sampling patterns: multirate sampling and random selection sampling. It is shown that those sampling patterns guarantee perfect recovery of the line spectra, and that the reconstruction can be achieved in a poly-logarithmic time with respect to the full observation case. Moreover, the sub-Nyquist recovery capabilities of such sampling patterns are highlighted. The atomic soft thresholding method is adapted in the presented framework to estimate sparse spectra in noisy environments, and a scalable algorithm for its resolution is proposed.

Index Terms

Sampling theory, line spectral estimation, super-resolution, sub-Nyquist sampling, multirate sampling, convex optimization, dimensionality reduction.

I. INTRODUCTION

Compressed sensing techniques have proven to be of great interests for detecting, estimating and denoising sparse signals lying on discrete spaces. On the practical side, the applications of sparse modeling are many: single molecule imaging via fluorescence, blind source separation in speech processing, precise separation of multiple celestial bodies in astronomy, or super-resolution radar, are among those. However, the discrete gridding required by the compressed sensing framework weaken the recovery performances, and more precisely the system *resolution*: the required minimal separation between two components of the sparse signal to be efficiently distinguished by an observation process.

In the recent years, a particular enthusiasm has been placed on solving sparse linear inverse problems over continuous dictionaries. This aims to recover the smallest finite subset of components generating a signal, and lying in a continuous space, by discrete observations of this signal distorted by a kernel function. Considering such approach raises new theoretical and practical concerns, in particular, those problems are commonly infinitely ill-posed.

This paper will discuss the spectral spikes recovery problem, also known as line spectrum estimation problem, which is probably one of the most fundamental and important illustration of sparse modeling over continuous spaces. For the spectral case, a complex time signal x is said to follow the s -spikes model if and only if it reads

$$\forall t \in \mathbb{R}, \quad x(t) = \sum_{r=1}^s \alpha_r e^{i2\pi \xi_r t}, \quad (\text{I.1})$$

whereby $\Xi = \{\xi_r, r \in \llbracket 1, s \rrbracket\}$ is the ordered set containing the s spectral components generating the signal x , and $\alpha = \{\alpha_r, r \in \llbracket 1, s \rrbracket\}$ the one of their associated complex amplitudes.

In the *total observation* framework, i.e. when observing $n \in \mathbb{N}$ uniform samples of the form $y[k] = x\left(\frac{k}{f}\right)$ for some sampling frequency $f \in \mathbb{R}^+$, the frequency estimation problem is naturally defined as building a consistent estimator $(\bar{\Xi}, \bar{\alpha})$ of the parameters (Ξ, α) , that are supposed to be unknown, of the time signal x based on the knowledge of $y \in \mathbb{C}^n$. This problem is obviously ill-posed, and since no assumption is a priori made on the number of frequencies s to estimate, there are infinitely many pairs $(\bar{\Xi}, \bar{\alpha})$ that are consistent with the observations. As for illustration purpose, the discrete Fourier transform of the observation vector y forms a consistent spectral representation of the signal x by n spectral spikes at locations $\bar{\xi}_k = \frac{k}{n}f$ in the frequency domain. However, this representation has generally no reason to be sparse, in the sense that a time signal x drawn from the s -spikes model will be represented by $n > s$ non-null spectral coefficients; unless the all the elements of Ξ exactly belongs to the spectral grid $\{\frac{k}{n}f, k \in \mathbb{Z}\}$.

Among all those consistent estimators, the one considered to be optimal, in the sparse recovery context, is the one returning the sparsest spectral distribution, i.e., the one that outputs a spectral support $\bar{\Xi}_0$ achieving the smallest cardinality \bar{s}_0 . Consequently, under total observation, by denoting \hat{x} the spectrum of x , the optimal spectral estimator \hat{x}_0 of \hat{x} can be written as the output of an optimization program of the form

$$\begin{aligned} \hat{x}_0 &= \arg \min_{\hat{x} \in D_1} \|\hat{x}\|_0 \\ \text{subject to} \quad & y = \mathcal{F}_{n,f}(\hat{x}), \end{aligned} \quad (\text{I.2})$$

where $\|\cdot\|_0$ represents the limit of the p pseudo-norm towards 0, counting the cardinality of the support. D_1 denotes the space of absolutely integrable spectral distributions, and $\mathcal{F}_{n,f}$ is the inverse discrete time Fourier transform for the sampling frequency $f \in \mathbb{R}^+$ defined by,

$$\mathcal{F}_{n,f} : D_1 \rightarrow \mathbb{C}^n \quad (\text{I.3})$$

$$\hat{x} \mapsto q : q[k] = \int_{\mathbb{R}} e^{i2\pi \frac{\xi}{f} k} d\hat{x}(\xi), \quad \forall k \in \llbracket 0, n-1 \rrbracket.$$

In case of absence of ambiguity on f , its notation will be simplified to \mathcal{F}_n .

Program (I.2) is non-convex, and the combinatorial nature of “ L_0 ” minimization leaves the direct formulation of this problem practically unsolvable. A commonly proposed workaround consists in analyzing the output of a relaxed problem, obtained by swapping the cardinality cost function $\|\cdot\|_0$ into a minimization of the total-variation norm over the spectral distribution domain $\|\cdot\|_{\text{TV}}$, defined by

$$\|\hat{x}\|_{\text{TV}} = \sup_{f \in C(\mathbb{R}), \|f\|_{\infty} \leq 1} \Re \left[\int_{\mathbb{R}} \overline{f(\xi)} d\hat{x}(\xi) \right],$$

where $C(\mathbb{R})$ denotes the space of continuous complex functions of the real variable. The total-variation norm can be interpreted as an extension of the L_1 norm to the distribution domain. This relaxation leads to the formulation of the convex program

$$\begin{aligned} \hat{x}_{\text{TV}} &= \arg \min_{\hat{x} \in D_1} \|\hat{x}\|_{\text{TV}} \\ \text{subject to } y &= \mathcal{F}_{n,f}(\hat{x}). \end{aligned} \tag{I.4}$$

Sufficient conditions for the tightness of this relaxation have been successfully addressed in [1], [2], [3], [4]: Problem (I.4) is known to output a spectral distribution \hat{x}_{TV} that is equal to the optimal solution \hat{x}_0 of the original Problem (I.2) under the mild separation assumption between the spikes in the frequency domain

$$\Delta_{\mathbb{T}} \left(\frac{1}{f} \Xi \right) \geq \frac{2.52}{n-1}, \tag{I.5}$$

provided that the number of measurements n is greater than some constant, and whereby $\Delta_{\mathbb{T}}(\cdot)$ is the set minimal warp around distance over the elementary torus $\mathbb{T} = [0, 1)$ defined by

$$\forall \Omega \subset \mathbb{R}, \quad \Delta_{\mathbb{T}}(\Omega) = \min \left\{ \text{frac}(\nu - \nu'), (\nu, \nu') \in \Omega^2, \nu \neq \nu' \right\},$$

and whereby $\text{frac}(\cdot)$ denotes the fractional part of any real number. Nevertheless, the estimate $\hat{x}_{\text{TV}} = \hat{x}_0$ will correspond to true distribution \hat{x} only if the Nyquist criteria is met, since an ambiguity modulo f stands in the spectral domain due to the aliasing effect generated by the uniform sampling process.

Related work on line spectral estimation

Up to recent years, most of the approaches to recover the off-the-grid spikes generating sparse signals were based on *subspace construction* methods. It is the case of the popular and proven MUSIC [5] and ESPRIT [6] methods, building tap delayed subspaces from the measurements, and making use of their low rank properties to locate the frequencies while denoising signals. A more recent method [7], based on annihilating and Cadzow filtering, describes an algebraic framework to estimate the set of continuous frequencies. If many of those methods have been shown to build consistent estimates, little is known about the theoretical spectral accuracy of those estimates under noisy observations.

The interest for approaching the line spectrum estimation problem under the lens of *convex optimization* has been increasing after that the recent work [1] established the optimality of convex relaxation under the previously

discussed conditions. It has been shown in [3] that such optimality still holds with high probability when extracting at random a small number of observations and discarding the rest of it.

The convex approach has been proven to be robust to noise, achieving near optimal mean-square error in Gaussian noise [8] under full measurements. The dispersion in L_1 norm in the time domain has been bounded in [2] for an arbitrary noise distribution. The sufficient separability criterion on the spikes has been enhanced in [4], and authors of [9] demonstrated that the estimated time signal converges in quadratic norm to the time signal x without any spectral separability conditions when the number of observations grows large.

The line spectral estimation problem is a practically important sub-case for the wider theory for pulse stream deconvolution. A general analysis of this framework is presented in [10], sufficient conditions of the tightness of the convex relaxation approach have been proposed [11], while [12] provides necessary ones. Authors of [13] proved that the deconvolution of spikes is possible without separation assumption for a broad class of distortion kernels, including the Gaussian one. On the computational side, several algorithms have been proposed to bridge the high computational cost of solving the relaxed Program (I.4), including a space discretization approach in [14], and an enhanced gradient search for sparse inverse problems in [15].

Many extensions of the spectral spikes model have been studied. The recent works [16], [17] extend to the case of multi-dimensional spikes, proving the efficiency of convex relaxations, although the resolution degrades with the order of the model. Estimation from multiple measurement vectors (MMV) has been proposed in [18], [19]. More generic models involving spectral deconvolution of spikes from unknown kernels have been studied in [20].

Other relaxation approaches to recover the spectral spikes exist in the literature. In [21], a nuclear norm minimization over the set of Hankel matrices were proved to return exact estimates without the need of any separation condition. Authors of [22] recently considered a relaxation using log-penalty functions achieving better empirical performances. However the robustness of those estimators to noisy environments remains unexplored.

Finally, on the practical side, the super-resolution theory of spikes has found application to super-resolution fluorescence microscopy and more recently to super-resolution radar imaging [23].

We emphasize on the fact that the cited studies address the line spectrum estimation problem under full observations $y \in \mathbb{C}^n$ acquired uniformly for some sampling frequency.

Focus and organization of this paper

If line spectral search is a theoretically promising approach to recover sparse spectra with very high precision, the computational complexity of the convex relaxation approach (II.2) remains the principal flaw to its use in practice. A direct approach to recover the spectra \hat{x}_{TV} using classic convex solvers grows as $\mathcal{O}(n^7)$ in the number of measurements n and becomes unrealistic when dealing with more than a few hundred of them.

This work aims to address the complexity issue by recovering the spectrum of the probed time signal x via *partial observations* $y \in \mathbb{C}^m$, obtained as linear combinations of the output of a uniform sampler $y_{\text{raw}} \in \mathbb{C}^n$, such that $y = My_{\text{raw}}$. The sub-sampling matrix $M \in \mathbb{C}^{m \times n}$ defines the linear combinations to apply on the raw output of the uniform sampler. We show that, under an unrestrictive admissibility condition on the sub-sampling

matrix M , the line spectral estimation problem can be reformulated as a semidefinite program of dimension $m + 1$. Moreover we study some categories of sub-sampling matrix and derive sufficient conditions for optimal recovery of the spectrum \hat{x} of the probed signal from sub-Nyquist sampling rates. We show that our approach can bring *orders of magnitude* changes to the computational complexity of the recovery, turning the standard polynomial time algorithm into equivalent ones of *poly-logarithmic* orders.

The present work is essentially organized in three parts. In the first part, Section II presents the *partial line spectral estimation* framework and states generic conditions for the recoverability of any time signal x following Model (I.1). It further introduces our main result in Theorem II.4, establishing the recoverability of x from the output of a semidefinite program of dimension $m + 1$. An explicit formulation of this program is provided for the remarkable case of so called selection matrices.

The second part of this work studies the recoverability of x from partial measurement acquired through a sub-sampling matrix M having a selection based structure. Two selection patterns are studied in details. The first one is presented in Section III, and treat the case where the output $y \in \mathbb{C}^m$ is generated by a multirate sampling systems: a system formed by a set of uniform samplers working at potentially different delays and frequencies. It is shown in Theorem III.5 that, under a common alignment property, involving certain conditions on the rates and the delays between the samplers, the output of relaxed approach to the line spectral estimation is tight. Furthermore, the sub-Nyquist recovery capabilities of the studied framework are highlighted, and the complexity gain of using such sampling model is discussed. Section IV presents the random selection sub-sampling model firstly introduced in [3] and shows that it can be used to reconstruct signal following the spikes model in a poly-logarithmic computational time.

In the last part of this paper, we address in Section V the estimation problem from noisy measurements by extending the atomic soft thresholding (AST) method proposed in [8] to our observation framework. A fast and scalable algorithm based on the alternative direction method of multipliers (ADMM) is presented in Section VI to estimate the spectral spikes from partial sampling. Finally, Section VII presents a detailed proof of Theorem II.4 that relies on an elegant extension of the Gram parametrization property of trigonometric polynomials to subspaces of polynomials.

II. DIMENSIONALITY REDUCTION FOR PARTIALLY OBSERVED SYSTEMS

A. Problem setup

We consider the estimation problem of a continuous time signal x following the spikes model (I.1) from m partial observations constructed linearly from the n ($n \geq m$) outputs of a uniform sampler. This sampler acquires the signal x uniformly at a given frequency $f \in \mathbb{R}^+$. The output of $y_{\text{raw}} \in \mathbb{C}^n$ of the sampler, before reduction, reads $y_{\text{raw}}[k] = x\left(\frac{k}{f}\right)$ for every sampling index $k \in \llbracket 0, n - 1 \rrbracket$. The observation vector $y \in \mathbb{C}^m$ is linked to the uniform acquisition y_{raw} by the linear relation $y = My_{\text{raw}}$ where $M \in \mathbb{C}^{m \times n}$ is the *sub-sampling matrix* of the system, which is assumed to be known.

As explained before, the line spectrum recovery problem consists in finding the continuous time signal x_0 that matches the observations y while having the sparsest spectral distribution \hat{x}_0 . In other terms, \hat{x}_0 has to be composed by the combination of spikes in the spectral domain of minimal cardinality s_0 . This “ L_0 ” minimization problem is called *partial line spectral estimation problem*, and can be described on an analogue manner to Program (I.2)

$$\begin{aligned} \hat{x}_{M,0} &= \arg \min_{\hat{x} \in D_1} \|\hat{x}\|_0 \\ \text{subject to} \quad &y = M\mathcal{F}_n(\hat{x}). \end{aligned} \tag{II.1}$$

The program is known to be NP-hard in the general case due to the combinatorial search imposed by the “ L_0 ” minimization. Therefore, we naturally introduce the total-variation counterpart to this problem in the same manner than (I.4), leading to

$$\begin{aligned} \hat{x}_{M,\text{TV}} &= \arg \min_{\hat{x} \in D_1} \|\hat{x}\|_{\text{TV}} \\ \text{subject to} \quad &y = M\mathcal{F}_n(\hat{x}). \end{aligned} \tag{II.2}$$

In the presented work, we address two fundamental issues arising from the formulation of the convex formulation (II.2):

- *Computational complexity*: Can one solve this convex problem in a computational time depending only on the dimension of the observations m ?
- *Recoverability*: Can one find sub-sampling matrices M guarantying the recoverability of the signal x (i.e. $\hat{x}_{M,\text{TV}} = \hat{x}_0$)?

B. Notations

We firstly introduce some notations that will be used in the rest of this work. For any complex number z , we write by \bar{z} its conjugate. The adjunction of \mathbf{X} is denoted \mathbf{X}^* , wherever \mathbf{X} is a vector, a matrix, or a linear operator. The transposition of a matrix or a vector \mathbf{X} is written \mathbf{X}^\top . If $P \in \mathbb{C}^{n-1}[X]$ is a complex polynomial of the form $P(z) = \sum_{k=0}^{n-1} p_k z^k$ then its conjugate is denoted P^* and verifies $P^*(z) = \sum_{k=0}^{n-1} \bar{p}_k z^k$ for all $z \in \mathbb{C}$. Unless stated differently, vectors of \mathbb{C}^n are indexed in $\llbracket 0, n-1 \rrbracket$ so that every vector $u \in \mathbb{C}^n$ writes $u = [u_0, \dots, u_{n-1}]^\top$. The space of square matrices and the one of Hermitian matrices of dimension n with complex coefficients are respectively denoted $M_n(\mathbb{C})$ and $S_n(\mathbb{C})$. The cone of positive Hermitian matrices of same dimension is denoted $S_n^+(\mathbb{C})$. Vectorial spaces of matrices are all endowed with the Frobenius inner product denoted $\langle \cdot, \cdot \rangle$ and defined by $\langle A, B \rangle = \text{tr}(A^* B)$, where $\text{tr}(\cdot)$ is the trace operator. The canonical Toeplitz Hermitian matrix generator in

dimension n , denoted \mathcal{T}_n , is defined by

$$\mathcal{T}_n : \mathbb{C}^n \rightarrow \mathbb{M}_n(\mathbb{C})$$

$$u \mapsto \mathcal{T}_n(u) = \begin{bmatrix} u_0 & u_1 & \dots & u_{n-1} \\ \overline{u_1} & u_0 & \dots & u_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{u_{n-1}} & \overline{u_{n-2}} & \dots & u_0 \end{bmatrix}. \quad (\text{II.3})$$

Its adjoint \mathcal{T}_n^* is characterized for every matrix $H \in \mathbb{M}_n(\mathbb{C})$ by

$$\forall k \in \llbracket 0, n-1 \rrbracket, \quad \mathcal{T}_n^*(H)[k] = \langle \Theta_k, H \rangle = \text{tr}(\Theta_k^* H),$$

whereby Θ_k is the elementary Toeplitz matrix equals to 1 on the k^{th} upper diagonal and zero elsewhere, i.e.

$$\forall (i, j) \in \llbracket 0, n-1 \rrbracket^2, \quad \Theta_k(i, j) = \begin{cases} 1 & \text{if } j - i = k \\ 0 & \text{otherwise.} \end{cases}$$

For every matrix $M \in \mathbb{C}^{m \times n}$, $m \leq n$, we denote by \mathcal{R}_M the operator given by

$$\mathcal{R}_M : \mathbb{C}^n \rightarrow \mathbb{M}_m(\mathbb{C})$$

$$u \mapsto \mathcal{R}_M(u) = M \mathcal{T}_n(u) M^*.$$

Its adjoint \mathcal{R}_M^* is consequently characterized for every matrix $S \in \mathbb{M}_m(\mathbb{C})$ by $\mathcal{R}_M^*(S) = \mathcal{T}_n^*(M^* S M)$.

A selection matrix $C_{\mathcal{I}} \in \{0, 1\}^{m \times n}$ for a subset $\mathcal{I} \subseteq \llbracket 0, n-1 \rrbracket$ of cardinality m is a boolean matrix whose rows are equal to $\{e_k^*, k \in \mathcal{I}\}$, where $e_k \in \mathbb{C}^n$ is the k^{th} vector of the canonical basis of \mathbb{C}^n . For a given subset \mathcal{I} , there are $m!$ possible associated sub-sampling matrices, all obtained by permutation of their rows. For readability, we reduce the respective notations of the operators $\mathcal{R}_{C_{\mathcal{I}}}$ and $\mathcal{R}_{C_{\mathcal{I}}}^*$ to $\mathcal{R}_{\mathcal{I}}$ and $\mathcal{R}_{\mathcal{I}}^*$ for such matrices.

C. Dual problem and certifiability

It has been shown in [3] that the primal problem (II.2) admits for Lagrange dual problem a certain semidefinite program when the sub-sampling matrix is selection matrix $C_{\mathcal{I}}$. This result easily extends in our context for any sub-sampling matrix M as stated by the following proposition.

Lemma II.1 (Dual characterization). *The dual feasible set \mathcal{D}_M of Problem (II.2) is characterized by*

$$\mathcal{D}_M = \left\{ c \in \mathbb{C}^m, \begin{cases} q = M^* c \\ \|Q(e^{i2\pi\nu})\|_{\infty} \leq 1 \end{cases} \right\},$$

whereby $Q \in \mathbb{C}^{n-1}[X]$ is the complex polynomial having for coefficients vector $q \in \mathbb{C}^n$. The Lagrangian dual of

Problem (II.2) takes the semidefinite form,

$$\begin{aligned}
 c_\star &= \arg \max_{c \in \mathbb{C}^m} \Re(y^\top c) \\
 \text{subject to } & \begin{bmatrix} H & q \\ q^* & 1 \end{bmatrix} \succeq 0 \\
 & \mathcal{T}_n^*(H) = e_0 \\
 & q = M^* c.
 \end{aligned} \tag{II.4}$$

Proposition II.2 (Dual certifiability). *If there exists a polynomial $Q_\star \in \mathbb{C}^{n-1}[X]$ having for coefficients vector $q_\star \in \mathbb{C}^n$ satisfying the conditions*

$$\begin{cases} q_\star \in \text{range}(M^*) \\ Q_\star(e^{i2\pi \frac{\xi_r}{f}}) = \text{sign}(\alpha_r), \quad \forall \xi_r \in \Xi \\ |Q_\star(e^{i2\pi\nu})| < 1, \quad \text{otherwise,} \end{cases} \tag{II.5}$$

then the solutions of the Programs (I.2), and (II.2) are unique and one has $\hat{x}_0 = \hat{x}_{M,TV}$. Moreover, $\hat{x} = \hat{x}_{M,TV}$ up to an aliasing factor modulo f .

Proof: Any polynomial Q_\star satisfying the last two interpolation conditions of (II.5) maximizes the dual of Problem (I.4) over the feasible set D_{I_n} , and qualifies as a dual certificate of the same problem. Thus, the solution of Program (I.4) is unique and satisfies $\hat{x}_0 = \hat{x}_{TV}$ [1]. By strong duality, the primal problem (I.4) and its dual reach the same optimal objective value, denoted κ_\star .

By the first condition of (II.5), $q_\star = M^* c_\star$ for some $c_\star \in \mathbb{C}^m$. Since $c \in D_M \Leftrightarrow M^* c \in D_{I_n}$ for all $c \in \mathbb{C}^m$, c_\star is dual optimal for the partial problem (II.2) and reaches the dual objective κ_\star . By strong duality, κ_\star also minimizes the primal objective of (II.2). Finally, every feasible point of (II.2) is feasible for (I.4). We conclude by uniqueness of \hat{x}_{TV} on the equality $\hat{x}_0 = \hat{x}_{TV} = \hat{x}_{M,TV}$. Finally $\hat{x} = \hat{x}_0$ (and thus $\hat{x}_{M,TV}$) up to an ambiguity modulo f is a direct consequence of Shannon's sampling theorem. ■

Any polynomial Q_\star satisfying the conditions (II.5) will be called *dual certificate* for the partial line spectral estimation problem. Finding meaningful sufficient conditions for the existence of such dual certificate is a difficult problem in the general case. One might expect their existence under two main conditions. The first one comes as a quite intuitive application of the principle stated in [12]: the spikes of the signal \hat{x} have to obey a minimal separability condition of the kind (I.5) (for a potentially different constant). The second one is on the sub-sampling matrix M , which has to somehow preserve the spectral properties of \hat{x} , and will be discussed latter.

Sufficient conditions for the existence of a dual certificate will be detailed in Section III and Section IV for two different classes of sub-sampling matrices. Generic results, valid for any arbitrary sub-sampling matrix M , are still lacking and remain an open area of research.

D. Main result

Lemma II.1 proposes to recover the spectral support of the time signal x by firstly solving a semidefinite program of dimension $n + 1$, and in the latter, to read its output c_* as a polynomial $Q_* \in \mathbb{C}^{n-1}[X]$, where $q_* = M^* c_*$. The spectral support of x is estimated by the points where this polynomial reaches 1 in modulus around the unit circle. However, this method is not satisfactory on a computational point of view. The complexity of the SDP (II.4) is driven by the size of its linear matrix inequality, here of size $n + 1$, while the *essential dimension* of partial recovery problem (II.2) is equal to the number of measurements $m \leq n$. In this section, it is shown that, if the matrix M admits a simple admissibility criterion, Program (II.4) is equivalent to another SDP involving a matrix inequality of lower dimension equal to $m + 1$.

Definition II.3 (Admissibility condition). A sub-sampling matrix $M \in \mathbb{C}^{m \times n}$ is said to be *admissible* if and only if M is full rank and $e_0 \in \text{range}(M^*)$, where $e_0 \in \mathbb{C}^n$ is the first vector of the canonical basis indexed in $\llbracket 0, n - 1 \rrbracket$.

Now we are ready to state the main result for this work, whose full demonstration is provided in Section VII.

Theorem II.4 (Dimensionality reduction). *If the sub-sampling matrix $M \in \mathbb{C}^{m \times n}$ is admissible, the Lagrange dual problem of Problem (II.2) is equivalent to the low-dimensional semidefinite program*

$$\begin{aligned} c_* &= \arg \max_{c \in \mathbb{C}^m} \Re(y^T c) \\ \text{subject to} \quad & \begin{bmatrix} S & c \\ c^* & 1 \end{bmatrix} \succeq 0 \\ & \mathcal{R}_M^*(S) = e_0. \end{aligned} \tag{II.6}$$

A few remarks are in order regarding the statement of Theorem II.4. First of all, the measurement matrix M has to be admissible for the theorem to hold. If this condition is not respected, the feasible set of SDP (II.6) is empty, and obviously differ from the dual feasible set \mathcal{D}_M . Secondly, the linear constraint $\mathcal{R}_M^*(S) = e_0$ has an explicit dimension that is still equal to n . However, since M is fixed and known, one can restrict this linear constraint to the span of $\mathcal{R}_M^*(S_m)$ which is of dimension lower than $\min \left\{ n, \frac{m(m+1)}{2} \right\} = \mathcal{O}(m^2)$. An explicit characterization of this constraint is provided in Section II-E when M is a selection matrix.

E. Case of selection sub-sampling matrices

Selection matrices constitutes a particularly interesting type of sub-sampling matrices, and arise in many practical applications. Their use is natural in signal processing occur when dealing with sampling models with missing entries. In this section, we highlight fundamental properties of the partial line spectrum estimation problem from selection based sub-sampling. We start by giving a direct characterization of the admissibility of a matrix $C_{\mathcal{I}}$.

Lemma II.5. *A selection matrix $C_{\mathcal{I}} \in \{0, 1\}^{m \times n}$ for a subset $\mathcal{I} \subseteq \llbracket 0, n - 1 \rrbracket$ for cardinality m is admissible in the sense of Definition II.3 if and only if $0 \in \mathcal{I}$.*

The proof of the above is trivial and arise directly from the definition of $C_{\mathcal{I}}$. The next proposition explicits the structure of $\mathcal{R}_{\mathcal{I}}^*$ and recast the linear constraint $\mathcal{R}_{\mathcal{I}}^*(S) = r$ under a more friendly set of equations.

Proposition II.6. *Let $\mathcal{I} \subset \llbracket 0, n-1 \rrbracket$ be a subset of cardinality m and consider any selection matrix $C_{\mathcal{I}} \in \mathbb{C}^{m \times n}$ for this subset. Define by \mathcal{J} the set of its pairwise differences $\mathcal{J} = \mathcal{I} - \mathcal{I}$, and by $\mathcal{J}_+ = \{j \in \mathcal{J}, j \geq 0\}$ its positive elements. There exists a skew-symmetric partition of the square $\llbracket 1, m \rrbracket^2$ into $p = |\mathcal{J}_+|$ subsets $\{J_k, k \in \mathcal{J}_+\}$ given by the support of the matrices $\{C_{\mathcal{I}}^* \Theta_k C_{\mathcal{I}}\}_{k \in \mathcal{J}_+}$ satisfying*

$$\begin{cases} J_k \cap J_l = \emptyset, & \forall (k, l) \in \mathcal{J}_+^2, k \neq l, \\ (i, j) \in \bigcup_{k \in \mathcal{J}_+} J_k \Leftrightarrow (j, i) \notin \bigcup_{k \in \mathcal{J}_+} J_k, & \forall (i, j) \in \llbracket 1, m \rrbracket^2, i \neq j, \\ (i, i) \in \bigcup_{k \in \mathcal{J}_+} J_k, & \forall i \in \llbracket 1, m \rrbracket, \end{cases}$$

such that,

$$\forall S \in S_m(\mathbb{C}), \mathcal{R}_{\mathcal{I}}^*(S) = \sum_{k \in \mathcal{J}_+} \left(\sum_{(l, r) \in J_k} S_{l, r} \right) e_k, \quad (\text{II.7})$$

where $e_k \in \mathbb{C}^n$ is the k^{th} vector of the canonical basis of \mathbb{C}^n indexed in $\llbracket 0, n-1 \rrbracket$.

Proof: Using the adjoint decomposition of the operator $\mathcal{R}_{\mathcal{I}}^*$ on the canonical basis one has,

$$\begin{aligned} \forall S \in S_m(\mathbb{C}), \quad \mathcal{R}_{\mathcal{I}}^*(S) &= \sum_{k=0}^{n-1} \langle \mathcal{R}_{\mathcal{I}}(e_k), S \rangle e_k \\ &= \sum_{k=0}^{n-1} \langle C_{\mathcal{I}} \Theta_k C_{\mathcal{I}}^*, S \rangle e_k. \end{aligned} \quad (\text{II.8})$$

Let by $M_k \in M_m(\mathbb{C})$ the matrix given by $M_k = C_{\mathcal{I}} \Theta_k C_{\mathcal{I}}^*$ for all $k \in \llbracket 0, n-1 \rrbracket$. It remains to show that the support of the matrices $\{M_k\}_{k \in \llbracket 0, n-1 \rrbracket}$ are forming the desired partition. The general term of matrix M_k , obtained by direct calculation, reads

$$\forall (i, j) \in \llbracket 1, m \rrbracket^2, \quad M_k(i, j) = \begin{cases} 1 & \text{if } \mathcal{I}[j] - \mathcal{I}[i] = k \\ 0 & \text{otherwise,} \end{cases} \quad (\text{II.9})$$

for all $k \in \llbracket 0, n-1 \rrbracket$, whereby $\mathcal{I}[j]$ represents the j^{th} element of the index set \mathcal{I} for the ordering induced by the matrix $C_{\mathcal{I}}$. The general term (II.9) ensures that,

$$\begin{cases} M_0(i, i) = 1, & \forall i \in \llbracket 1, m \rrbracket \\ \sum_{k=0}^n M_k(i, j) = 1 \Leftrightarrow \sum_{k=0}^n M_k(j, i) = 0, & \forall (i, j) \in \llbracket 1, m \rrbracket^2, i \neq j, \\ k \notin \mathcal{J}_+ \Leftrightarrow M_k = 0_m, & \forall k \in \llbracket 0, n-1 \rrbracket, \end{cases}$$

where 0_m is the null element of $M_m(\mathbb{C})$. Since the matrices $\{M_k\}_{k \in \llbracket 0, n-1 \rrbracket}$ are constituted of boolean entries, the two first assertions yields the set of supports $\{J_k\}_{k \in \llbracket 0, n-1 \rrbracket}$ of $\{M_k\}_{k \in \llbracket 0, n-1 \rrbracket}$ forms an skew-symmetric partition of $\llbracket 1, m \rrbracket^2$. The third one states that only $p = |\mathcal{J}_+|$ elements of this partition are non-trivial. After removing those

null matrices, the set $\{J_k\}_{k \in \mathcal{J}_+}$ remains a partition of $\llbracket 1, m \rrbracket^2$. We conclude using Equation (II.8) that,

$$\begin{aligned} \forall S \in S_m(\mathbb{C}), \quad \mathcal{R}_{\mathcal{I}}^*(S) &= \sum_{k \in \mathcal{J}_+} \langle M_k, S \rangle e_k \\ &= \sum_{k \in \mathcal{J}_+} \left(\sum_{(l,r) \in J_k} S_{l,r} \right) e_k. \end{aligned}$$

■

This proposition highlights several major properties of the equation $\mathcal{R}_{\mathcal{I}}^*(S) = r$ for $r \in \mathbb{C}^n$:

- The linear equation is solvable if and only if r is supported in \mathcal{J}_+ , and since $M_0 = I_m$, $r_0 \in \mathbb{R}$.
- If so, the equation is equivalent to solve $p = |\mathcal{J}_+|$ linear forms. Those p forms are independent one from the other in the sense that they are acting on disjoint extractions of the matrix S .
- The order of each of those forms is smaller than m , i.e., each form involves at most m terms of S .
- The total number of unknowns appearing in this system is exactly $\frac{m(m+1)}{2}$.

In Section VI, a highly scalable algorithm to solve the SDP (II.6) for selection matrices, taking advantage of the hereby presented properties, will be presented.

III. MULTIRATE SAMPLING SYSTEMS

A. Observation model

A multirate sampling system (MRSS) on a continuous time signal x is defined by a set \mathbb{A} of p distinct grids (or samplers) \mathcal{A}_j , $j \in \llbracket 1, p \rrbracket$. Each grid is assimilated to a triplet $\mathcal{A}_j = (f_j, \gamma_j, n_j)$, where $f_j \in \mathbb{R}^+$ is its sampling frequency, $\gamma_j \in \mathbb{R}$ is its processing delay, expressed in sample unit for normalization purposes, and $n_j \in \mathbb{N}$ the number of measurements acquired by the grid. We assume those intrinsic characteristics to be known. The output $y_j \in \mathbb{C}^{n_j}$ of the grid \mathcal{A}_j sampling a complex time signal x following the s -spikes model (I.1) reads

$$\forall k \in \llbracket 0, n_j - 1 \rrbracket, \quad y_j[k] = \sum_{r=1}^s \alpha_r e^{i2\pi \frac{\xi_r}{f_j} (k - \gamma_j)}. \quad (\text{III.1})$$

Applications of the MRSS framework are numerous in signal processing. It occurs when sampling in parallel the output of a common channel in order to get benefits from cleverly designed sampling frequencies and delays; such design appears, for example, in modern digitalization with variable bit-rates and analysis of video and audio streams. The MRSS framework is also naturally fitted to describe sampling processes in distributed sensor networks: each node, with limited processing capabilities, samples at its own rate, a delayed version of a complex signal. Collected data are then sent and merged at a higher level processing unit, performing a global estimation of the spectral distribution on a joint manner.

The frequency estimation problem consists, as explained earlier, in finding the sparsest spectral density that jointly matches the p observation vectors y_j for all $j \in \llbracket 1, p \rrbracket$. Equivalently to (I.2), this problem can be presented by a

combinatorial minimization program of the L_0 pseudo-norm over the set of spectral distributions:

$$\begin{aligned} \hat{x}_0 &= \arg \min_{\hat{x} \in D_1} \|\hat{x}\|_0 \\ \text{subject to } y_j &= \mathcal{L}_j(\hat{x}), \quad \forall j \in \llbracket 1, p \rrbracket, \end{aligned} \quad (\text{III.2})$$

where \mathcal{L}_j is the linear operator denoting the effect of the spectral density on the samples acquired by the grid \mathcal{A}_j given by

$$\forall j \in \llbracket 1, p \rrbracket, \quad \mathcal{L}_j = \mathcal{F}_{n, f_j} \circ \mathcal{M}_{f_j}^{\gamma_j}, \quad (\text{III.3})$$

whereby the operator \mathcal{M}_τ , $\tau \in \mathbb{R}$ denotes the temporal shift (or spectral modulation) operator defined for all $h \in D_1$ by $\mathcal{M}_\tau(h)(\xi) = e^{-i2\pi\tau\xi}h(\xi)$ for all $\xi \in \mathbb{R}$.

Finally, it is important to notice that two different grids \mathcal{A}_j and $\mathcal{A}_{j'}$ may sample a value of the signal x at the same time instant on the respective sampling indexes k and k' , enforcing a relation of the kind $y_j[k] = y_{j'}[k']$. In the following we denote by $\tilde{m} = \sum_{j=1}^p n_j$ the total number of samples acquired by the system \mathbb{A} , and by $m \leq \tilde{m}$ the net number of observations, obtained after removing such sampling overlaps, so that m is the number of *independent observation constraints* of the system. The joint measurement vector is denoted $\tilde{y} = [y_1^T, \dots, y_p^T]^T \in \mathbb{C}^{\tilde{m}}$. We let by $y \in \mathbb{C}^m$ its net counterpart by discarding the redundancies of \tilde{y} , so that $y = P_{\mathbb{A}}\tilde{y}$ for some selection matrix $P_{\mathbb{A}} \in \{0, 1\}^{m \times \tilde{m}}$. The joint linear measurement constraint of Problem (III.2) can then be reformulated $y = \mathcal{L}(\hat{x})$, where the operator $\mathcal{L} \in (D_1 \mapsto \mathbb{C}^m)$ admits the partial operators $\{\mathcal{L}_j\}_{j \in \llbracket 1, p \rrbracket}$ as restrictions on the p subspaces induced by the construction of the net observation vector y .

B. Common grid expansion and SDP formulation

It is been shown in Section II that the dual problem can take the form of a low dimensional SDP whenever the observation operator \mathcal{L} can be written under the form $\mathcal{L} = M\mathcal{F}_n$ for some measurement matrix $M \in \mathbb{C}^{m \times n}$ satisfying the admissibility condition II.3. As highlighted in the proof of Lemma II.5, this remarkable property is due to the polynomial nature of the adjoint measurement operator \mathcal{L}^* . However, in the MRSS context, the dual observation operator defined by $\mathcal{L}^*(c) = \sum_{j=1}^m \mathcal{L}_j^*(c_j)$ does not take such polynomial form in the general case. A direct calculation reveals that $\mathcal{L}^*(c)$ is instead an exponential polynomial¹ for all $c \in \mathbb{C}^m$. Up to our knowledge, there is no welcoming algebraic characterization for optimization purposes of the dual feasible set $\mathcal{D}_{\mathbb{A}} = \{c \in \mathbb{C}^m, \|\mathcal{L}^*(c)\|_\infty \leq 1\}$. Therefore, the theory developed in Section II cannot be directly transcribed in the MRSS framework.

To bridge this concern, we restrict our analysis to the case where the observation operator admits a factorization of the form $\mathcal{L} = M\mathcal{F}_n$ for some $n \in \mathbb{N}$ and $M \in \mathbb{C}^{m \times n}$. The following aims to provide an algebraic criterion on the parameters $\{(f_j, \gamma_j, n_j)\}$ of \mathbb{A} for this hypothesis to hold. We will see that this extra hypothesis consists in supposing that the samples acquired by \mathbb{A} can be virtually aligned at a higher rate on another grid \mathcal{A}_+ . Such grid will be called common supporting grid for \mathbb{A} , and are defined as follows.

¹A function f of the complex variable z of the form $f(z) = \sum_{k=1}^m c_k z^{\gamma_k}$ for some $\{\gamma_k\}_{\llbracket 1, m \rrbracket} \subset \mathbb{R}$.

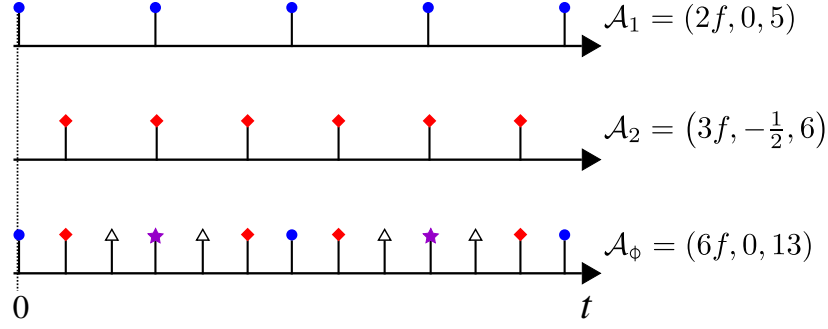


Figure III.1. A representation of a multirate sampling system \mathbb{A} composed of two arrays $(\mathcal{A}_1, \mathcal{A}_2)$, and its associated minimal common grid \mathcal{A}_ϕ . Purple stars in the common grid correspond to time instant acquired multiple times by the system \mathbb{A} , and blank triangles to omitted samples. In this example, the dimension of the minimal common grid is $n_\phi = 13$. The total number of observation of \mathbb{A} , $\tilde{m} = 5 + 6 = 11$, and the net number of observations is $m = 9$. Finally the equivalent observation set of the common grid is $\mathcal{I} = \{0, 1, 3, 5, 6, 7, 9, 11, 12\}$.

Definition III.1. A grid $\mathcal{A}_+ = (f_+, \gamma_+, n_+)$ is said to be a *common supporting grid* for a set of sampling grids $\mathbb{A} = \{\mathcal{A}_j\}_{j \in \llbracket 1, p \rrbracket}$ if and only if the set of samples acquired by the MRSS induced by \mathbb{A} is a subset of the one acquired by \mathcal{A}_+ . In formal terms, the definition is equivalent to,

$$\left\{ \frac{1}{f_j} (k_j - \gamma_j), j \in \llbracket 1, p \rrbracket, k_j \in \llbracket 0, n_j - 1 \rrbracket \right\} \subseteq \left\{ \frac{1}{f_+} (k - \gamma_+), k \in \llbracket 0, n_+ - 1 \rrbracket \right\}. \quad (\text{III.4})$$

The set of common supporting grids of \mathbb{A} is denoted by $\mathcal{C}(\mathbb{A})$. Moreover, a common supporting grid $\mathcal{A}_\phi = (f_\phi, \gamma_\phi, n_\phi)$ for \mathbb{A} is said to be *minimal* if and only it satisfies the minimality condition,

$$\forall \mathcal{A}_+ \in \mathcal{C}(\mathbb{A}), \quad n_\phi \leq n_+.$$

Finally, the *equivalent observation set* of the minimal common grid \mathcal{A}_ϕ , denoted by \mathcal{I} , is the subset of $\llbracket 0, n_\phi - 1 \rrbracket$ of cardinality m , formed by the k 's for which the time instant $\frac{1}{f_\phi} (k - \gamma_\phi)$ is acquired by \mathbb{A} .

It is clear that if $\mathcal{C}(\mathbb{A})$ is not empty then the minimal common supporting grid for \mathbb{A} exists and is unique. For ease of understanding, Figure III.1 illustrates the notion of common supporting grid by showing a MRSS formed by two arrays and their minimal common grid. Proposition III.2 states necessary and sufficient conditions in terms of the parameters of \mathbb{A} such that the set $\mathcal{C}(\mathbb{A})$ is not empty. The proof of this proposition is technical and delayed to Appendix D for readability.

Proposition III.2. Given a set of p grids $\mathbb{A} = \{\mathcal{A}_j = (f_j, \gamma_j, n_j)\}_{j \in \llbracket 1, p \rrbracket}$, the set $\mathcal{C}(\mathbb{A})$ is not empty if and only if there exist $f_+ \in \mathbb{R}^+$, $\gamma_+ \in \mathbb{R}$, a set of p positive integers $\{l_j\} \in \mathbb{N}^p$, and a set of p integers $\{a_j\} \in \mathbb{Z}^p$ satisfying $f_+ = l_j f_j$ and $\gamma_+ = l_j \gamma_j - a_j$ for all $j \in \llbracket 1, p \rrbracket$. Moreover a common grid $\mathcal{A}_\phi = (f_\phi, \gamma_\phi, n_\phi)$ is minimal, if and only if

$$\begin{cases} \gcd(\{a_j\}_{j \in \llbracket 1, p \rrbracket} \cup \{l_j\}_{j \in \llbracket 1, p \rrbracket}) = 1 \\ \gamma_\phi = \max_{j \in \llbracket 1, p \rrbracket} \{l_j \gamma_j\} \\ n_\phi = \max_{j \in \llbracket 1, p \rrbracket} \{l_j (n_j - 1) - a_j\}. \end{cases}$$

Remark III.3. Although the conditions of Proposition III.2 appear to be strong since one get $\mathcal{C}(\mathbb{A}) = \emptyset$ almost surely in the Lebesgue sense when the sampling frequencies and delays are drawn at random, assuming the existence of a common supporting grid for \mathbb{A} is not meaningless in our context. By density, one can approximately align the system \mathbb{A} on an arbitrary fine grid \mathcal{A}_ε , for any given maximal jitter $\varepsilon > 0$, and perform the proposed super-resolution on this common grid. The resulting error from this approximation can be interpreted as a “basis mismatch”. The detailed analysis of this approach will not be covered in this work, however, similar approximations can be found in the literature for the analogue atomic norm minimization view of the super-resolution problem [9]. We claim that those results extend in our settings and that the approximation error vanishes in the noiseless settings when going to the limit $\varepsilon \rightarrow 0$.

The next proposition concludes that the requested factorization of the linear observation operator \mathcal{L} is possible whenever $\mathcal{C}(\mathbb{A}) \neq \emptyset$.

Proposition III.4. *Let $\mathbb{A} = \{\mathcal{A}_j = (f_j, \gamma_j, n_j)\}_{j \in \llbracket 1, p \rrbracket}$ be a set of p arrays. The set $\mathcal{C}(\mathbb{A})$ is not empty if and only if there exists a subset $\mathcal{I} \subseteq [0, n_\phi - 1]$ of cardinality m such that the linear operator \mathcal{L} defining the equality constraint of the primal Problem (I.4) reads,*

$$\mathcal{L} = C_{\mathcal{I}} \left(\mathcal{F}_{n_\phi, f_\phi} \circ \mathcal{M}_{\frac{\gamma_\phi}{f_\phi}} \right),$$

whereby $\mathcal{A}_\phi = (f_\phi, \gamma_\phi, n_\phi)$ denotes the minimal grid of \mathbb{A} and where $C_{\mathcal{I}} \in \{0, 1\}^{m \times n_\phi}$ is a selection matrix of the subset \mathcal{I} . Moreover the sub-sampling matrix $C_{\mathcal{I}}$ is admissible in the sense of Definition II.3.

The proof of this proposition is detailed in Appendix B-B. The temporal translation $\mathcal{M}_{\frac{\gamma_\phi}{f_\phi}}$ has little impact in the analysis since a time domain shift leaves unchanged the spectral support of the probed signal x . One can consider the surrogate signal $x^\#(\cdot) = x\left(\cdot - \frac{\gamma_\phi}{f_\phi}\right)$, so that $\hat{x}^\# = \mathcal{M}_{\frac{\gamma_\phi}{f_\phi}}(\hat{x})$ and solve the line spectral estimation problem (II.2) for the linear constraint $\mathcal{L}^\# = C_{\mathcal{I}} \mathcal{F}_{n_\phi, f_\phi}$ via the reduction studied in Section II. The complex amplitudes of the spectral \hat{x} can be recover from its surrogate spectrum by the simple relation $e^{i2\pi \frac{\gamma_\phi}{f_\phi} \xi} \alpha^\#(\xi) = \alpha(\xi)$ for all $\xi \in \mathbb{R}$.

C. Dual certifiability and sub-Nyquist guarantees

In this section, sufficient conditions are presented to ensure that the conditions of Proposition II.2 are fulfilled. Those conditions guarantee the tightness of the total-variation relaxation and the optimality and uniqueness of the recovery $\hat{x}_0 = \hat{x}_{C_{\mathcal{I}}, \text{TV}}$. In addition to this result, it provides mild conditions to ensure a sub-Nyquist recovery of the spectral spikes at a rate f_ϕ from measurements taken at various lower rates $\{f_j\}_{j \in \llbracket 1, p \rrbracket}$. The proof of this result, presented in Appendix C, relies on previous polynomial construction methods presented in [1], [3], [9].

Theorem III.5. *Let $\mathbb{A} = \{\mathcal{A}_j = (f_j, \gamma_j, n_j)\}_{j \in \llbracket 1, p \rrbracket}$ be a set of sampling arrays. Suppose that $\mathcal{C}(\mathbb{A})$ is not empty, and denote by $\mathcal{A}_\phi = (f_\phi, \gamma_\phi, n_\phi)$ the minimal common supporting grid of \mathbb{A} . Assume that the system induced by \mathbb{A} satisfies at least one of the two following separability conditions,*

- Strong condition:

$$\forall j \in \llbracket 1, p \rrbracket, \quad \begin{cases} \Delta_{\mathbb{T}}\left(\frac{1}{f_j}\Xi\right) \geq \frac{2.52}{n_j-1} \\ n_j > 2000, \end{cases}$$

- Weak condition:

$$\exists j \in \llbracket 1, p \rrbracket, \quad \begin{cases} \Delta_{\mathbb{T}}\left(\frac{1}{f_j}\Xi\right) \geq \frac{2.52}{n_j-1} \\ n_j > 2000 \\ m \geq (l_j + 1)s, \end{cases}$$

then there exists a polynomial Q_* verifying the conditions (II.5) Proposition II.2. Consequently, $\hat{x}_0 = \hat{x}_{C_{\mathcal{I}}, TV}$. Moreover, $\hat{x} = \hat{x}_{C_{\mathcal{I}}, TV}$ up to an aliasing factor modulo f_ϕ .

Remark III.6. First of all, under the weaker proviso $n_j > 256$, the above results still hold in both cases when Ξ satisfies the more restrictive separability criterion $\Delta_{\mathbb{T}}\left(\frac{1}{f_j}\Xi\right) \geq \frac{4}{n_j-1}$.

The strong condition for Theorem III.5 is restrictive and does not particularly highlight any benefits from jointly estimating the spectral support compared to merging the p spectral estimates obtained by simple individual estimation at each sampler. However, the weak condition guarantees that frequencies of the time signal x can be recovered with an ambiguity modulo f_ϕ when jointly resolving the MRSS, while individual estimations would guarantee to recover them with an ambiguity modulo $f_j \leq f_\phi$. The weak condition requires standard spectral separation from a single array \mathcal{A}_j , and sufficient net measurements m of the time signal. The extra measurements $m - n_j$ corresponding to the other grids are *not uniformly aligned* with the sampler \mathcal{A}_j . Therefore the sampling system induced by \mathbb{A} achieves sub-Nyquist spectral recovery of the spectral spikes, and pushes away the classic spectral range f_j from a factor $\frac{f_\phi}{f_j} = l_j$. Nevertheless, the provided construction of the dual certificate results in a polynomial having a modulus close to unity on the aliasing frequencies induced by the zero forcing upscaling from f_j to f_ϕ . Consequently, one can expect to obtain degraded performances in noisy environments when the sub-sampling factor l_j becomes large.

D. Benefits of multirate measurements

Multirate sampling has been applied in many problematics arising from signal processing and telecommunications in order to reduce either the number of required measurements or the processing complexity [24]. There are three major benefits of making use of MRSS acquisition in the line spectral estimation problem. One might just think MRSS has an obvious way of increasing the number of samples acquired by system compared to a single grid measurement $\mathcal{A}_j \in \mathbb{A}$. This naturally leads to an enhanced *noise robustness*. More importantly, MRSS acquisition brings benefits in terms *spectral range* extension, and *spectral resolution* improvement. The spectral range extension (or sub-Nyquist) capabilities have been described in Theorem II.4. The spectral resolution — the minimal distance on the torus between two spectral spikes to guarantee their recovery —, is also expected to be enhanced in MRSS acquisition due to the observation of delayed versions of the time signal x , which virtually enlarges the global observation window. The resolution guarantees in MRSS will not be covered in this work and are left for future research.

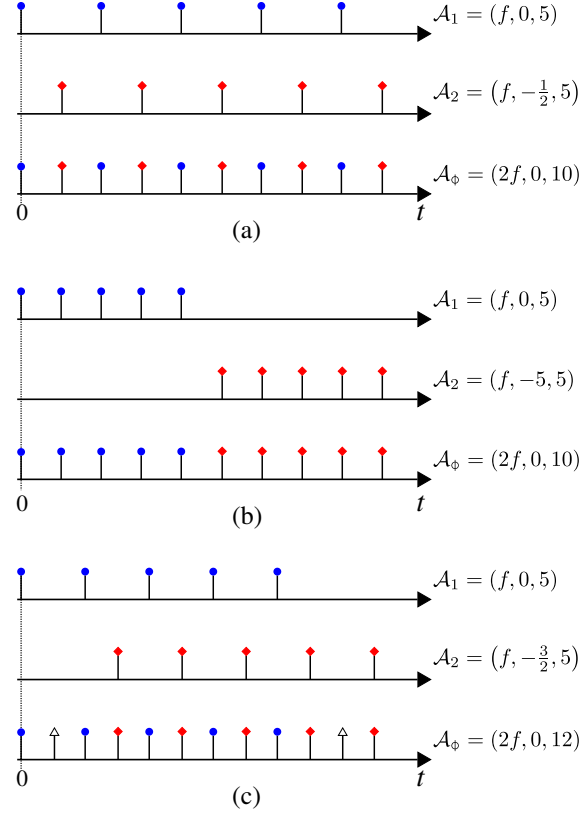


Figure III.2. A representation of three delay-only MRSS in different remarkable settings. In (a), the delay between the two samplers is exactly of half-unit, resulting in a doubled frequency range in the joint analysis. In (b), this delay is such that the overall process equivalently acquires samples on a doubled time frame, resulting in a doubled spectral resolution. Sub-figure (c) represents an hybrid case where both resolution improvement and spectral range extension are expected.

For the sake of clarity, Figure III.2 proposes a comprehensive illustration of the trade-off between range extension and resolution improvement for a delay-only MRSS constituted of two samplers \mathcal{A}_1 and \mathcal{A}_2 . In Figure III.2 (a), the delay between the two samplers is such that the joint uniform grid \mathcal{A}_Φ has no missing observations with a double sampling frequency. One trivially expects to recover the spikes location of x with aliasing ambiguity modulo $2f$. In III.2(b), the delay of \mathcal{A}_2 is set such that the resulting minimal common grid has a doubled observation window. \mathcal{A}_Φ fits again in the uniform observation framework analyzed in [1], and the sufficient spectral separation from the joint measurements is twice smaller than for the single estimation case. Finally a hybrid case is presented in Figure III.2(c), where one expect to get some spectral range and resolution improvements from a joint recovery approach.

E. Complexity improvements

Proposition III.4 states that, under the existence of a common grid, the selection operator $C_{\mathcal{I}} \in \{0, 1\}^{m \times n_\Phi}$ is admissible, consequently Theorem II.4 apply and the dual line spectrum estimation problem can be formulated, in the MRSS context, by an SDP of dimension $m + 1$. In this section, we highlight the important impact in term of

complexity in the MRSS case.

The original semidefinite program (II.4) involves a linear matrix inequality of dimension of $n_\phi + 1$. The actual value of n_ϕ , fully determined of the observation pattern induced by \mathbb{A} , reads

$$n_\phi = \max_{j \in \llbracket 1, p \rrbracket} \{l_j (n_j - 1) - a_j\},$$

whereby the parameters $\{(a_j, l_j)\}_{j \in \llbracket 1, p \rrbracket}$ are defined in Proposition III.2. This is particularly disappointing since n_ϕ grows at a speed driven by the product of the n_j 's, whereas the *essential dimension* m of the problem is given by the number of net observations acquired by the grid $m \leq \tilde{m} = \sum_{j=1}^p n_j$. We study the asymptotic ratio $\frac{m}{n_\phi}$ when the number grids p grows large in two different idealized instances of MRSS to illustrate that the reduced SDP formulation (II.6) brings *orders of magnitude* changes to the computational complexity of the line spectral estimation problem.

Suppose a delay-only MRSS, where \mathbb{A} is constituted of p grids given by $\mathcal{A}_1 = (f, 0, n_0)$ and $\mathcal{A}_j = \left(f, -\frac{1}{b_j}, n_0\right)$ for all $j \in \llbracket 2, p \rrbracket$. Moreover suppose that $\{b_j\}_{j \in \llbracket 2, p \rrbracket}$ are jointly coprime. It is easy to verify the $\mathcal{C}(\mathbb{A})$ is not empty in those settings, and that the minimal common grid \mathcal{A}_ϕ is given by $\mathcal{A}_\phi = \left(\left(\prod_{j=2}^p b_j\right) f, 0, \left(\prod_{j=2}^p b_j\right) n_0\right)$. One has $n_\phi = \Omega(b^p n_0)$ for some constant $b \in \mathbb{R}^+$, while $m = p n_0$. The ratio $\frac{m}{n_\phi} = o\left(\frac{p}{b^p}\right)$ and tends to 0 exponentially fast with the number of samplers m of the system.

On the other hand, suppose a synchronous coprime sampling system between the time instants 0 and T , where $\mathcal{A}_j = (k_j f, 0, k_j f T)$ for all $j \in \llbracket 1, p \rrbracket$ with $\gcd\{k_j, j \in \llbracket 1, p \rrbracket\} = 1$. Once again $\mathcal{C}(\mathbb{A})$ is not empty, and the minimal grid is characterized by the parameters $\mathcal{A}_\phi = \left(\left(\prod_{j=1}^p k_j\right) f, 0, \left(\prod_{j=1}^p k_j\right) f T\right)$. Consequently $\frac{m}{n_\phi} = \frac{\sum_{j=1}^p k_j}{\prod_{j=1}^p k_j}$ decreases in $o(k^{-p})$ for judicious choice of $\{k_j\}_{j \in \llbracket 1, p \rrbracket}$.

IV. RANDOM SELECTION SAMPLING

A. Observation model and previous results

In this section, we consider the line spectrum estimation problem from a category of selection matrix $C_{\mathcal{I}} \in \{0, 1\}^{m \times n}$ obtained by randomly selecting the observation subset \mathcal{I} . This problem has been introduced in [3], and sufficient conditions to guarantee the tightness of Program (II.2) have been provided. We hereby summarize those results and introduce our low dimensional approach to recover the frequencies of a sparse signal x in those measurement settings.

The observation subset $\mathcal{I} \subseteq \llbracket 0, n-1 \rrbracket$ is constructed by keeping at random, and independently from the others, each of the elements of $\llbracket 0, n-1 \rrbracket$ with probability p , and discarding the rest of it. As a result, \mathcal{I} has an expected cardinality $\bar{m} = \mathbb{E}[\|\mathcal{I}\|] = pn$. We consider a subset \mathcal{I} , of cardinality m resulting from the described stochastic process, and recall the following result from [3, Theorem I.1].

Theorem IV.1 (Tang, Bhaskar, Shah, Recht '12). *Consider the partial observation problem (II.2) with a sub-sampling matrix $M = C_{\mathcal{I}} \in \{0, 1\}^{m \times n}$ drawn according to the random selection sampling model. Suppose that the observed signal x following model (I.1) satisfies the spectral separability condition $\Delta_{\mathbb{T}}\left(\frac{1}{f}\Xi\right) \geq \frac{4}{n-1}$. Moreover,*

suppose that the phases of the complex amplitudes $\{\alpha_r\}_{r \in \llbracket 1, s \rrbracket}$ characterizing the signal x are drawn independently and uniformly at random in $[0, 2\pi)$. Consider any positive number $\delta > 0$. There exists a constant $C > 0$ such that if

$$m \geq C \max \left\{ \log^2 \frac{n}{\delta}, s \log \frac{s}{\delta} \log \frac{n}{\delta} \right\},$$

then there exists, with probability greater than $1 - \delta$, a polynomial Q_\star verifying the conditions (II.5).

Consequently, the output of the relaxed Problem (II.2) is unique and verifies $\hat{x}_0 = \hat{x}_{C_{\mathcal{I}}, TV}$. Moreover, $\hat{x} = \hat{x}_{C_{\mathcal{I}}, TV}$ up to an aliasing factor modulo f .

B. Dimensionality reduction

The dimensionality reduction result presented in Theorem II.4 requires an admissible selection matrix $C_{\mathcal{I}}$, i.e. that $0 \in \mathcal{I}$. In the latter, we show that we can always fall back into this case via some simple considerations similar to the one described in Section III-B. Let $k_0 = \min \mathcal{I}$, and let $x^\sharp(\cdot) = x\left(\cdot - \frac{k_0}{f}\right)$. In the spectral domain the definition reads $\hat{x}^\sharp = \mathcal{M}_{\frac{k_0}{f}} \hat{x}$, and the line spectral estimation problem can be equivalently solved for the spectral density \hat{x}^\sharp for the measurement constraint

$$\mathcal{L}^\sharp = C_{\mathcal{I}-k_0} \mathcal{F}_n.$$

One has $0 \in \mathcal{I} - k_0$ and the selection matrix $C_{\mathcal{I}-k_0} \in \mathbb{C}^{m \times n}$ is thus admissible in the sense of Definition II.3. It is therefore possible to recover \hat{x}^\sharp from the reduced SDP (II.6), and to reconstruct in a second time \hat{x} via the simple phase shift $\alpha(\xi) = e^{-i2\pi \frac{\gamma_\Phi}{f_\Phi} \xi} \alpha^\sharp(\xi)$. We are now allowed to conclude on the following result.

Corollary IV.2. *Under the same hypothesis than Theorem IV.1, the reduced SDP (II.6) of dimension $m+1$ outputs a polynomial Q_\star verifying the conditions (II.5).*

Theorem IV.1 guarantees a high-probability recovery of supporting frequencies of the probed signal whenever the number of measurement grow essentially as the logarithm of n . Therefore using the reduced SDP (II.6) to solve the line spectral estimation problem brings again *orders of magnitude* changes in term of computational complexity. The complexity is lowered from solving the SDP of dimension n , to solving a SDP having a *poly-logarithmic* dimension dependency $m = \mathcal{O}\left(\max\left\{\log^2 \frac{n}{\delta}, s \log \frac{s}{\delta} \log \frac{n}{\delta}\right\}\right)$.

V. SPECTRAL ESTIMATION IN NOISE

Up to here, only the case of noise-free spectral estimation has been studied. In this part, we consider partial noisy observations of a sparse signal x following the spikes model given in (I.1) under the form

$$\begin{cases} y_{\text{raw}}[k] = \sum_{r=1}^s \alpha_r e^{i2\pi \frac{\xi_r}{f} k} + w[k], & \forall k \in \llbracket 0, n-1 \rrbracket \\ y = M y_{\text{raw}}, \end{cases}$$

for some sub-sampling matrix $M \in \mathbb{C}^{m \times n}$. The noise vector $w \in \mathbb{C}^n$ is assumed to be drawn according to the spherical n dimensional complex Gaussian distribution $\mathcal{N}(0, \sigma^2 I_n)$. We introduce an adapted version of the original

Atomic Soft Thresholding (AST) method, introduced in [8] to denoise the spectrum of x and attempt to retrieve the set of frequencies Ξ supporting the spectral spikes. The AST method is reviewed to perform in the partial observation context. Its Lagrange dual version is introduced, and benefits from the same dimensionality reduction properties than discussed in Section II. The Primal-AST problem consists in optimizing the cost function

$$\hat{x}_{\text{TV}} = \arg \min_{\hat{x} \in D_1} \|\hat{x}\|_{\text{TV}} + \frac{\tau}{2} \|y - M\mathcal{F}_n(\hat{x})\|_2^2, \quad (\text{V.1})$$

whereby $\tau \geq 0$ is a regularization parameter trading between the sparsity of the recovered spectrum and the denoising power. Making use of Proposition II.6, if M satisfies the admissibility condition given in Definition II.3, the Dual-AST problem is equivalent to the low-dimensional semidefinite program

$$\begin{aligned} c_\star &= \arg \max_{c \in \mathbb{C}^m} \Re(y^\top c) - \frac{\tau}{2} \|c\|_2^2 \\ \text{subject to} \quad & \begin{bmatrix} S & c \\ c^* & 1 \end{bmatrix} \succeq 0 \\ & \mathcal{R}_M^*(S) = e_0. \end{aligned} \quad (\text{V.2})$$

Slatter's condition holds once again for Problem (V.1), and strong duality between (V.1) and (V.2) is ensured. Applying the results in [9], the choice of regularization parameter $\tau = \gamma\sigma\sqrt{m\log m}$, for some $\gamma > 1$, is suitable to guarantee a perfect asymptotic recovery of the spectral distribution \hat{x} , while providing accelerated rates of convergence.

VI. ESTIMATION VIA ALTERNATING DIRECTION METHOD OF MULTIPLIERS

A. Interior point methods and ADMM

Computing the solution of semidefinite program using out of the box SDP solvers such as SUDEMI [25] or SDPT3 [26] requires at most $\mathcal{O}\left((m_{\text{lin}}^2 + m_{\text{lin}})^{3.5}\right)$ operations where m_{lin} is the dimension of the linear matrix inequality, and m_{lin} the dimension of the linear constraints. For the dual-AST program (V.2), $m_{\text{lin}} = m + 1$ and $m_{\text{lin}} \leq \frac{m(m+1)}{2}$, and approaching the optimal dual solution will cost $\mathcal{O}(m^7)$ operations using those interior point methods. It appears to be unrealistic to recover the sparse line spectrum of x that way when the number of observations exceeds a few hundreds.

In the same spirit than in [8], we derive the steps and update equations to approach the optimal solution via the alternating direction method of multipliers (ADMM). Unlike the original work, we choose to perform ADMM on the dual space instead of the primal one, and adjust the update steps in order to take advantage of the low dimensionality of (V.2). The overall idea of this algorithm is to cut the augmented Lagrangian of the problem into a sum of separable sub-functions. Each iteration consists in performing independent local minimization on each of those quantities. The interested reader can find a detailed survey of this method in [27].

We restrict our analysis to the case of partially observed systems where the sub-matrix is a selection matrix $C_{\mathcal{I}} \in \{0, 1\}^{m \times n}$ for some subset $\mathcal{I} \subseteq \llbracket 0, n-1 \rrbracket$ of cardinality m . We will see that the properties of such matrices detailed in Section II-E will help breaking down the iterative steps of dual ADMM on an elegant manner. Before any further analysis, the Dual-AST (V.2) has to be restated into a more friendly form to derive the ADMM update equations. In our approach, we propose the following augmented formulation

$$\begin{aligned} c_{\star} &= \arg \min_{c \in \mathbb{C}^m} -\Re(y^{\top} c) + \frac{\tau}{2} \|c\|_2^2 \\ \text{subject to } \quad Z &\succeq 0 \\ Z &= \begin{bmatrix} S & c \\ c^* & 1 \end{bmatrix} \\ \sum_{(i,j) \in J_k} S_{i,j} &= \delta_k, \quad k \in \mathcal{J}_+, \end{aligned} \tag{VI.1}$$

whereby δ_k is the Kronecker symbol. It is immediate, using Proposition II.6, to verify that Problems (V.2) and (VI.1) are actually equivalent.

B. Lagrangian separability

We denote by L the restricted Lagrangian of the Problem (VI.1), obtained by ignoring the semidefinite constraint $Z \succeq 0$. In order to ensure plain differentiability with respect to the variables S and Z , ADMM seeks to minimize an augmented version L_+ of L , with respect to the semidefinite inequality constraint that was put apart. This augmented Lagrangian L_+ is introduced as follows

$$L_+(Z, S, c, \Lambda, \mu) = L(Z, S, c, \Lambda, \mu) + \frac{\rho}{2} \left\| Z - \begin{bmatrix} S & c \\ c^* & 1 \end{bmatrix} \right\|_F^2 + \frac{\rho}{2} \sum_{k \in \mathcal{J}_+} \left(\sum_{(i,j) \in J_k} S_{i,j} - \delta_k \right)^2,$$

whereby the variables $\Lambda \in S_{m+1}(\mathbb{C})$ and $\mu \in \mathbb{C}^{|\mathcal{J}_+|}$ denote respectively the Lagrange multipliers associated with the first and the second equality constraints of Problem (VI.1). The regularizing parameter $\rho > 0$ is set to ensure a well conditioned differentiability and to fasten the convergence speed of the alternating minimization towards the global optimum of the cost function L_+ . For clarity and convenience, the following decompositions of the parameters Z and Λ are introduced

$$Z = \begin{bmatrix} Z_0 & z \\ z^* & \zeta \end{bmatrix} \quad \Lambda = \begin{bmatrix} \Lambda_0 & \lambda \\ \lambda^* & \eta \end{bmatrix}.$$

Moreover, for any square matrix $A \in M_m(\mathbb{C})$, we let by $A_{J_k} \in \mathbb{C}^{|J_k|}$ the vector constituted of the terms $\{A_{i,j}, (i,j) \in J_k\}$. The order in which the elements of J_k are extracted and placed in this vector has no importance, as long as, once chosen, it remains the same for every matrix A . This allows to decompose the augmented Lagrangian into

$$L_+(Z, S, c, \Lambda, \mu) = L_c(z, c, \lambda) + L_\gamma(\zeta, \eta) + \sum_{k \in \mathcal{J}_+} L_k(Z_{0,J_k}, S_{J_k}, \Lambda_{0,J_k}),$$

whereby each of the sub-functions reads

$$\begin{aligned}
L_c(z, c, \lambda) &= -\Re(y^\top c) + \frac{\tau}{2} \|c\|_2^2 + 2 \langle \lambda, z - c \rangle + \rho \|z - c\|_2^2 \\
L_\gamma(\zeta, \eta) &= \langle \eta, \zeta - 1 \rangle + \frac{\rho}{2} (\zeta - 1)^2 \\
\forall k \in \mathcal{J}_+, \quad L_k(Z_{0,J_k}, S_{J_k}, \Lambda_{0,J_k}) &= \langle \Lambda_{0,J_k}, Z_{0,J_k} - S_{J_k} \rangle + \mu_k \left(\sum_{(i,j) \in J_k} S_{i,j} - \delta_k \right) \\
&\quad + \frac{\rho}{2} \|Z_{0,J_k} - S_{J_k}\|_2^2 + \frac{\rho}{2} \left(\sum_{(i,j) \in J_k} S_{i,j} - \delta_k \right)^2.
\end{aligned}$$

C. Update rules

The ADMM will consist in successively performing the following decoupled update steps:

$$\begin{aligned}
c^{t+1} &\leftarrow \arg \min_c L_c(z^t, c, \lambda^t) \\
\forall k \in \mathcal{J}_+, \quad S_{J_k}^{t+1} &\leftarrow \arg \min_{S_{J_k}} L_k(Z_{0,J_k}^t, S_{J_k}, \Lambda_{0,J_k}^t) \\
S_{j,i}^{t+1} &\leftarrow \overline{S_{i,j}^{t+1}}, \quad \forall (i,j) \in \bigcup_{k \in \mathcal{J}_+} J_k \\
Z^{t+1} &\leftarrow \arg \min_{Z \succeq 0} L_+(Z, S^{t+1}, c^{t+1}, \Lambda^t, \mu^t) \\
\Lambda^{t+1} &\leftarrow \Lambda^t + \rho \left(Z^{t+1} - \begin{bmatrix} S^{t+1} & c^{t+1} \\ c^{t+1*} & 1 \end{bmatrix} \right) \\
\forall k \in \mathcal{J}_+, \quad \mu^{t+1}(k) &\leftarrow \mu^t(k) + \rho \left(\sum_{(i,j) \in J_k} S_{i,j}^{t+1} - \delta_k \right).
\end{aligned}$$

Since the linear constraint $\mathcal{R}_{\mathcal{I}}^*(S) = e_0$ has an effect limited to the subspace $\{\mathcal{R}_{\mathcal{I}}(e_k)\}_{k \in \mathcal{J}_+}$, the third update step is necessary to maintain the Hermitian structure of the matrix S^{t+1} at every iteration. The update steps for the variables c^{t+1} and $\{S_{J_k}^{t+1}\}_{k \in \mathcal{J}_+}$ are performed at each iteration by canceling the gradient of their partial augmented Lagrangian and admit, in the presented settings, closed form expressions given by

$$\begin{aligned}
c^{t+1} &= \frac{1}{2\rho + \tau} (\bar{y} + 2\rho z^t + 2\lambda^t) \\
\forall k \in \mathcal{J}_+, \quad S_{J_k}^{t+1} &= \left(Z_0^t + \frac{1}{\rho} \Lambda_0^t \right)_{J_k} - \left(\sum_{(i,j) \in J_k} \left(Z_0^t + \frac{\Lambda_0^t}{\rho} \right)_{i,j} - \left(\delta_k - \frac{\mu_k^t}{\rho} \right) \right) j_{|J_k|}
\end{aligned}$$

whereby $\bar{y} \in \mathbb{C}^m$ denotes the conjugate of the observation vector y , and j_v is the all-one vector of \mathbb{C}^v for all $v \in \mathbb{N}$.

The update Z^{t+1} reads at the t^{th} iteration

$$\begin{aligned}
Z^{t+1} &\in \arg \min_{Z \succeq 0} \|Z - Y^t\|_F^2 \\
Y^t &= \begin{bmatrix} S^{t+1} & c^{t+1} \\ c^{t+1*} & 1 \end{bmatrix} - \frac{\Lambda^t}{\rho},
\end{aligned}$$

which can be interpreted as an orthogonal projection of Y^t onto $S_{m+1}^+(\mathbb{C})$ for the Frobenius inner product. This projection can be computed by looking for the eigenpairs of Y^t , and setting all negative eigenvalues to 0. More precisely, denoting $Y^t = V^t D^t V^{t*}$ an eigen-decomposition of Y^t , one get $Z^{t+1} = V^t D_+^t V^{t*}$ where D_+^t is a diagonal matrix whose j^{th} diagonal entry $d_+^t[j]$ satisfies $d_+^t[j] = \max\{d^t[j], 0\}$.

D. Computational complexity

On the computational point of view, at each step of ADMM, the update c^{t+1} is a vector addition and performed in a linear time $\mathcal{O}(m)$. On every extractions $S_{j_k}^{t+1}$ of S^{t+1} , the update equation is assimilated to a vector averaging requiring $\mathcal{O}(|J_k|)$ operations when firstly calculating the common second term of the addition. Since $\bigcup_{k \in \mathcal{J}_+} J_k = \frac{m(m+1)}{2}$, we conclude that the global update of the matrix S^{t+1} is done in $\mathcal{O}(m^2)$. The update of Z^{t+1} requires the computation of its spectrum, which can be done in $\mathcal{O}(m^3)$ via power method. Finally updating the multipliers Λ^{t+1} and μ^{t+1} consist in simple matrix and vector additions, thus of order $\mathcal{O}(m^2)$.

To summarize, the projection is the most costly operation of the loop. Each step of ADMM method runs in $\mathcal{O}(m^3)$ operations, which is a significant improvement compared to the infeasible path approached used by SDP solvers requiring around $\mathcal{O}(m^7)$ operations.

VII. PROOF OF THEOREM II.4

A. Gram parametrization of trigonometric polynomials

We start the demonstration by introducing a couple of notations and by a brief review of the Gram parametrization theory of trigonometric polynomials. For every non-zero complex number $z \in \mathbb{C}^*$, its n -length power vector $\psi_n(z) \in \mathbb{C}^n$ is defined by $\psi_n(z) = [1, z, \dots, z^{n-1}]^T$. A complex trigonometric polynomial $R \in \mathbb{C}^{\bar{n}}[X]$ of order $\bar{n} = 2n - 1$ is a linear combination of complex monomials with positive and negative exponents absolutely bounded by n . Such polynomial R reads

$$\forall z \in \mathbb{C}^*, \quad R(z) = \sum_{k=-n+1}^{n-1} r_k z^k.$$

It is easy to verify that a complex trigonometric polynomial takes real values around the unit circle, i.e. $R(e^{i\theta}) \in \mathbb{R}$ for all $\theta \in [0, 2\pi)$, if and only if vector $r \in \mathbb{C}^{\bar{n}}$ satisfies the Hermitian symmetry condition

$$\forall k \in \llbracket 0, n-1 \rrbracket, \quad r_{-k} = \overline{r_k}. \quad (\text{VII.1})$$

Every element of $\mathbb{C}^{\bar{n}}[X]$ can be associated with a subset of $M_n(\mathbb{C})$, called Gram set, as defined bellow.

Definition VII.1. A complex matrix $G \in M_n(\mathbb{C})$ is a *Gram matrix* associated with the trigonometric polynomial R if and only if

$$\forall z \in \mathbb{C}^*, \quad R(z) = \psi_n(z^{-1})^T G \psi_n(z).$$

Such parametrization is, in general, not unique and we denote by $\mathcal{G}(R)$ the set of matrices satisfying the above relation. $\mathcal{G}(R)$ is called *Gram set* of R .

The next proposition characterizes the Gram set of a complex trigonometric polynomial taking real values on the unit circle via a simple linear relation.

Proposition VII.2. *Let $R \in \mathbb{C}^{\bar{n}}[X]$ if a complex trigonometric polynomial taking real values around the unit circle. Let $G \in M_n(\mathbb{C})$, then $G \in \mathcal{G}(R)$ if and only if the relation*

$$\mathcal{T}_n^*(G) = r$$

holds, where $r = [r_0, \dots, r_{n-1}]^T \in \mathbb{C}^n$ is the vector containing the coefficients of R corresponding to its positive exponents.

The interested reader is invited to refer to [28, Theorem 2.3] for a proof and further consequences of this proposition.

B. Compact representations of polynomials in subspaces

The notion of Gram sets adapts to every complex trigonometric polynomial; if R is of order \bar{n} , it defines a subset $\mathcal{G}(R)$ of matrices from $M_n(\mathbb{C})$. In our context, the polynomials of interest have to belong to a low dimensional subspace characterized by the sub-sampling matrix $M \in \mathbb{C}^{m \times n}$. Finding compact Gram representations, involving matrices of lower dimensions, is of crucial interest for reflecting the low dimensionality of Problem (II.6). In the following, Definition VII.3 introduces the notion of compact representations, and Corollary VII.4 derives an immediate characterization of those when the considered polynomial takes real values around the unit circle.

Definition VII.3. A complex trigonometric polynomial $R \in \mathbb{C}^{\bar{n}}[X]$ is said to admit a *compact Gram representation* on a matrix $M \in \mathbb{C}^{m \times n}$, $m \leq n$ if and only if there exists a matrix $G \in M_m(\mathbb{C})$ such that the relation

$$\begin{aligned} \forall z \in \mathbb{C}^*, \quad R(z) &= \psi_n(z^{-1})^T M^* G M \psi_n(z) \\ &= \phi_M(z^{-1})^T G \phi_M(z) \end{aligned}$$

holds, where $\phi_M(z) = M^* \psi_n(z)$. We denote by $\mathcal{G}_M(R)$ the subset of complex matrices satisfying this property.

Corollary VII.4. *Let $R \in \mathbb{C}^{\bar{n}}[X]$ be a complex trigonometric polynomial taking real values around the unit circle. Let $G \in M_n(\mathbb{C})$, then $G \in \mathcal{G}_M(R)$ if and only if the relation*

$$\mathcal{R}_M^*(G) = r$$

holds, where $r = [r_0, \dots, r_{n-1}]^T \in \mathbb{C}^n$ is the vector containing the coefficients of R corresponding to its positive exponents.

The proof of this corollary is a direct consequence of Proposition VII.2 and of the definition of \mathcal{R}_M^* given in Section II-B.

C. Bounded real lemma for polynomial subspaces

This part aims to demonstrate a novel result, synthesized in Theorem VII.6, giving a low-dimensional semidefinite equivalence of the condition $|Q(e^{i2\pi\nu})| \leq |P(e^{i2\pi\nu})|$ for all $\nu \in \mathbb{T}$ when P and Q are complex polynomials whose respective coefficients vectors $p, q \in \mathbb{C}^n$ lie in the range of a linear operator M^* , where $M \in \mathbb{C}^{m \times n}$. Before going into its statement, it is necessary to introduce the intermediate Proposition VII.5 which highlights the compatibility between canonical partial order relations defined on the set trigonometric polynomials and the one of Hermitian matrices.

Proposition VII.5. *Let $R \in \mathbb{C}^{\bar{n}}[X]$ and $R' \in \mathbb{C}^{\bar{n}}[X]$ be two complex trigonometric polynomials taking real values around the unit circle. Let by $M \in \mathbb{C}^{m \times n}$ a full rank matrix and suppose that the sets $\mathcal{G}_M(R)$ and $\mathcal{G}_M(R')$ are both non-empty. Then the inequality $R'(e^{i2\pi\nu}) \leq R(e^{i2\pi\nu})$ holds for all $\nu \in \mathbb{T}$ if and only if for every two Hermitian matrices $G \in \mathcal{G}_M(R)$ and $G' \in \mathcal{G}_M(R')$, one has $G' \preceq G$.*

The proof of Proposition VII.5 is provided in Appendix A. We are now able to state and demonstrate a generic algebra result, linking the dominance around the unit circle of polynomials belonging to some subspace of $\mathbb{C}^{n-1}[X]$ with an Hermitian semidefinite inequality. Theorem VII.6 plays a key role in the demonstration of Theorem II.4.

Theorem VII.6 (Constrained Bounded Real Lemma). *Let P and Q be two polynomials of $\mathbb{C}^{n-1}[X]$ with respective coefficients vectors $p, q \in \mathbb{C}^n$. Moreover, suppose that p and q belong to the range of M^* , where $M \in \mathbb{C}^{m \times n}$ is a full rank matrix, and denote by $u \in \mathbb{C}^m$ a vector satisfying $q = M^*u$. Define by R the trigonometric polynomial $R(z) = P(z^{-1})P^*(z)$ for all $z \in \mathbb{C}^*$, and call $r \in \mathbb{C}^n$ its positive coefficients such that R can be written under the form $R(z) = r_0 + \sum_{k=1}^{n-1} (r_k z^k + \bar{r}_k z^{-k})$ for all $z \in \mathbb{C}^*$. Then the inequality*

$$\forall \nu \in \mathbb{T}, \quad |Q(e^{i2\pi\nu})| \leq |P(e^{i2\pi\nu})|$$

holds if and only if there exists a matrix $S \in \mathcal{S}_m(\mathbb{C})$ verifying

$$\left\{ \begin{array}{l} \begin{bmatrix} S & u \\ u^* & 1 \end{bmatrix} \succeq 0 \\ \mathcal{R}_M^*(S) = r. \end{array} \right.$$

Proof: Denote by R' the trigonometric polynomial defined by $R'(z) = Q(z^{-1})Q^*(z)$ for all $z \in \mathbb{C}^*$. Since the identities $R'(e^{i2\pi\nu}) = |Q(e^{-i2\pi\nu})|^2$ and $R(e^{i2\pi\nu}) = |P(e^{-i2\pi\nu})|^2$ are verified for all $\nu \in \mathbb{T}$, the inequality $|Q(e^{i2\pi\nu})| \leq |P(e^{i2\pi\nu})|$ is equivalent to $R'(e^{i2\pi\nu}) \leq R(e^{i2\pi\nu})$ for all $\nu \in \mathbb{T}$. In the latter, we derive conditions for this second inequality to hold.

First of all, since p is the range of M^* , one can find a vector $v \in \mathbb{C}^m$ verifying $p = M^*v$. It comes that

$$\begin{aligned} \forall \nu \in \mathbb{T}, \quad R(e^{i2\pi\nu}) &= P(e^{-i2\pi\nu}) P^*(e^{i2\pi\nu}) \\ &= \psi_n(e^{i2\pi\nu})^* p p^* \psi_n(e^{i2\pi\nu}) \\ &= \psi_n(e^{i2\pi\nu})^* M^* v v^* M \psi_n(e^{i2\pi\nu}). \end{aligned}$$

Thus, the rank one matrix vv^* belongs to $\mathcal{G}_M(R)$. On a similar manner, one has $uu^* \in \mathcal{G}_M(R')$ and the sets $\mathcal{G}_M(R)$ and $\mathcal{G}_M(R')$ are non-empty. The conditions of Proposition VII.5 are met. Consequently, the inequality $R'(e^{i2\pi\nu}) \leq R(e^{i2\pi\nu})$ holds for all $\nu \in \mathbb{T}$ if and only if there exists a Hermitian matrix $S \in \mathcal{G}_{\mathcal{I}}(R)$ satisfying $S \succeq uu^*$. By Corollary VII.4, $S \in \mathcal{G}_{\mathcal{I}}(R)$ is equivalent to $\mathcal{R}_M^*(S) = r$. Moreover, making use of the Schur complement, one has

$$S \succeq uu^* \Leftrightarrow \begin{bmatrix} S & u \\ u^* & 1 \end{bmatrix} \succeq 0,$$

which concludes on the desired result. ■

D. Proof of the main statement

We conclude in this section by proving that the dual SDP (II.4) is equivalent to a compact one (II.6) whenever the sub-sampling operator $M \in \mathbb{C}^{m \times n}$ is admissible in the sense of Definition II.3. The proof of this result is a consequence of the constraint bounded real lemma presented in the previous Section VII-C.

Proof: By Lemma II.1 the dual feasible set \mathcal{D}_M of the relaxed problem (II.2) writes

$$\mathcal{D}_M = \left\{ c \in \mathbb{C}^m, \begin{bmatrix} q = M^*c \\ \|Q(e^{i2\pi\nu})\|_{\infty} \leq 1 \end{bmatrix} \right\},$$

where $Q \in \mathbb{C}^{n-1}[X]$ is the polynomial having for coefficients vector $q \in \mathbb{C}^n$. The core idea of the proof consist in recasting the inequality on the infinite norm of Q by

$$\forall \nu \in \mathbb{T}, \quad |Q(e^{i2\pi\nu})| \leq |P_1(e^{i2\pi\nu})|,$$

where P_1 is the constant unitary polynomial of $\mathbb{C}^{n-1}[X]$. Define by R_1 the constant complex trigonometric polynomial of $\mathbb{C}^{\bar{n}}[X]$ reading $R_1(z) = P_1(z^{-1}) P_1^*(z) = 1$ for all $z \in \mathbb{C}^*$. The vector $r_1 \in \mathbb{C}^n$ of its positive monomial coefficients writes $r_1 = e_0$.

For any $c \in \mathcal{D}_M$, the vector $q = M^*c$ belongs to the range of M^* . Moreover, since M is admissible, $r_1 = e_0 \in \text{range}(M^*)$. The condition of application of Theorem VII.6 are met, and the equivalence

$$c \in \mathcal{D}_M \Leftrightarrow \exists S \text{ Hermitian s.t. } \begin{cases} \begin{bmatrix} S & c \\ c^* & 1 \end{bmatrix} \succeq 0 \\ \mathcal{R}_M^*(S) = e_0 \end{cases}$$

holds, which concludes the demonstration. ■

VIII. DISCUSSION AND FUTURE WORK

The construction of dual polynomial Q_\star matching the conditions (II.5) has been successfully achieved in the partial observation case only for very specific categories of sub-sampling matrices. It would be of great interest to characterize more finely the conditions for the existence of such polynomials in the generic case. In particular, highlighting the loss of resolution induced by the choice of the sub-sampling matrix M can have an impact in understanding the trade-off between the heavy high resolution recovery provided by the full measurement framework, and the fast coarser estimate provided by partial sub-sampling. However, such approaches would require to restrict the construction of the dual polynomial proposed in [1] to subspaces formed by non-aligned observations, which can be technically challenging.

Finally, we suggested in Remark III.3 that ε -approximating common grid could be used as an approximation when the conditions of Proposition III.2 do not strictly hold, and proposed to consider their performances under the lens of an analogue basis mismatch problem. Since the dimensionality of the reduced SDP (II.6) recovering the frequencies does not depend on the size of the common grid, one can wonder how the proofs presented in this paper can extend to a super-resolution theory of sparse spectrum from fully asynchronous measurements by letting the observation operator $\mathcal{F}_{kn,kf}$ deviating when k grows large.

APPENDIX A

PROOF OF PROPOSITION VII.5

We start the demonstration by proving the following lemma.

Lemma A.1. *Let $R \in \mathbb{C}^{\bar{n}}[X]$ be a complex trigonometric polynomial. Let $M \in \mathbb{C}^{m \times n}$, $m \leq n$ be a full rank matrix, and suppose that $\mathcal{G}_M(R)$ is not empty. The following assertions hold:*

R takes real values around the unit circle if and only if $\mathcal{G}_M(R)$ intersects the set of Hermitian matrices, i.e.

$$\forall \nu \in \mathbb{T}, \quad R(e^{i2\pi\nu}) \in \mathbb{R} \Leftrightarrow \mathcal{G}_M(R) \cap S_m(\mathbb{C}) \neq \emptyset. \quad (\text{A.1})$$

R takes positive values around on the unit circle if and only if $\mathcal{G}_M(R)$ intersects the cone of positive Hermitian matrices, i.e.

$$\forall \nu \in \mathbb{T}, \quad R(e^{i2\pi\nu}) \in \mathbb{R}^+ \Leftrightarrow \mathcal{G}_M(R) \cap S_m^+(\mathbb{C}) \neq \emptyset, \quad (\text{A.2})$$

and every Hermitian matrix in $\mathcal{G}_T(R)$ is positive.

Proof: We start the demonstration by showing that the set $\mathcal{G}_M(R)$ is a convex set. The proof is immediate by taking any two matrices G and G' in $\mathcal{G}_M(R)$ and any real $\beta \in [0, 1]$. Recalling the definition of the compact Gram set, it yields

$$\begin{aligned} \forall z \in \mathbb{C}^*, \quad \phi_M(z^{-1})^\top (\beta G + (1 - \beta) G') \phi_M(z) &= \beta R(z) + (1 - \beta) R(z) \\ &= R(z). \end{aligned}$$

Thus $\beta G + (1 - \beta) G' \in \mathcal{G}_T(R)$, and the convexity follows.

We carry on the demonstration of Assertion (A.1) by showing that R takes real values around the unit circle if and only if the set $\mathcal{G}_T(R)$ is stable by Hermitian transposition. First of all, it is easy to see via Definition VII.1 of the set $\mathcal{G}_M(R)$ that

$$G \in \mathcal{G}_M(R) \Leftrightarrow G^* \in \mathcal{G}_M(R^*).$$

Moreover since R takes real values around the unit circle, its coefficient vector satisfies the symmetry property (VII.1) (and reciprocally), which translate into $\mathcal{G}_M(R) = \mathcal{G}_M(R^*)$. Combining the last two relations lead to the equivalence with the stability of $\mathcal{G}_M(R)$ by Hermitian transposition, i.e.

$$R(e^{i2\pi\nu}) \in \mathbb{R} \Leftrightarrow \forall G \in \mathcal{G}_M(R), G^* \in \mathcal{G}_M(R). \quad (\text{A.3})$$

We conclude the demonstration of Assertion (A.1) by taking any element $G \in \mathcal{G}_M(R)$, and by noticing

$$R(e^{i2\pi\nu}) \in \mathbb{R} \Leftrightarrow \forall G \in \mathcal{G}_M(R), \frac{G + G^*}{2} \in \mathcal{G}_M(R),$$

using the convexity and the stability of $\mathcal{G}_M(R)$ by Hermitian transposition. Since $\mathcal{G}_M(R)$ is not empty by assumption, it intersects non-trivially the set of Hermitian matrices.

Suppose now that R takes real positive values over the unit circle. Let by S a Hermitian matrix belonging to $\mathcal{G}_M(R)$ (Assertion (A.1) attests the existence of such matrix). It comes

$$\begin{aligned} \forall \nu \in \mathbb{T}, \quad R(e^{i2\pi\nu}) &= \psi_n(e^{-i2\pi\nu})^T M^* S M \psi_n(e^{i2\pi\nu}) \\ &= \phi_M(e^{-i2\pi\nu})^* S \phi_M(e^{i2\pi\nu}). \end{aligned}$$

Since the sub-sampling matrix M is full rank, the set $\{\phi_M(e^{i2\pi\nu}), \nu \in \mathbb{T}\}$ spans the whole vectorial space \mathbb{C}^m . Thus, the positivity of R is equivalent to the positivity of the Hermitian matrix S , concluding on the second statement of the lemma. \blacksquare

We are now ready to start the demonstration the Proposition VII.5.

Denote respectively by $r, r' \in \mathbb{C}^n$ the respective positive coefficients of the trigonometric polynomials R and R' . The sets $\mathcal{G}_T(R)$ and $\mathcal{G}_T(R')$ are non-empty by assumption, and Lemma A.1 guarantees the existence of two Hermitian matrices S_0 and S'_0 belonging respectively to $\mathcal{G}_T(R)$ and $\mathcal{G}_T(R')$. Define by T the trigonometric polynomial

$$\begin{aligned} \forall \nu \in \mathbb{T}, \quad T(e^{i2\pi\nu}) &= R(e^{i2\pi\nu}) - R'(e^{i2\pi\nu}) \\ &= \phi_M(e^{i2\pi\nu})^* (S_0 - S'_0) \phi_M(e^{i2\pi\nu}). \end{aligned} \quad (\text{A.4})$$

Proving that R is greater than R' around the unit circle is equivalent to prove the positivity of T on the same domain. It is clear that the matrix $S_0 - S'_0$ belongs to $\mathcal{G}_M(T)$ and thus $\mathcal{G}_M(T)$ is not empty. By application of Lemma A.1, T is positive if and only if every Hermitian matrix H in the set $\mathcal{G}_M(T)$ is positive. We conclude that T is positive if and only if for every pair of Hermitian matrices $(S, S') \in \mathcal{G}_T(R) \times \mathcal{G}_T(R')$ one has $S \succeq S'$. \blacksquare

APPENDIX B

DUAL CHARACTERIZATION LEMMA II.1 AND PROOF OF PROPOSITION III.4

A. Proof of Lemma II.1

A standard Lagrangian analysis leads to a dual of (II.2) of the form

$$\begin{aligned} c_\star &= \arg \max_{c \in \mathbb{C}^m} \Re(y^\top c) \\ \text{subject to} \quad & \|\mathcal{F}_n^*(M^*c)\|_\infty \leq 1 \\ & q = M^*c. \end{aligned} \tag{B.1}$$

By direct calculation, one has

$$\begin{aligned} \forall c \in \mathbb{C}^m, \forall \xi \in \mathbb{R}, \quad \mathcal{F}_n^*(M^*c)(\xi) &= \mathcal{F}_n^*(q)(\xi) \\ &= \sum_{k \in \mathcal{I}} q_k e^{-i2\pi k \frac{\xi}{f}} \\ &= Q\left(e^{-i2\pi \frac{\xi}{f}}\right) \end{aligned}$$

where $q = M^*c$ is the coefficients vector of the polynomial $Q \in \mathbb{C}^{n-1}[X]$. The characterization of \mathcal{D}_M follows by noticing the invariance of the infinite norm over the transform $\xi \leftarrow -\xi$. The equivalence between Program (B.1) and an SDP is a direct consequence of the relation

$$\|Q(e^{i2\pi\nu})\|_\infty \leq 1 \Leftrightarrow \exists H \text{ Hermitian s.t. } \begin{cases} \begin{bmatrix} H & q \\ q^* & 1 \end{bmatrix} \succeq 0 \\ \mathcal{T}_n^*(H) = e_0. \end{cases}$$

A proof of this last assertion can be found in [28, Corollary 4.25]. ■

B. Proof of Proposition III.4

Proof: We recall from Equation (III.3) that for all $\hat{x} \in D_1$, one has,

$$\forall j \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 0, n_j - 1 \rrbracket, \quad \mathcal{L}_j[k] = \int_{\mathbb{R}} e^{i2\pi \frac{\xi}{f_j}(k - \gamma_j)} d\hat{x}(\xi).$$

Suppose that $\mathcal{C}(\mathbb{A})$ is not empty, the minimal common supporting grid $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ for \mathbb{A} exists. It comes by Equation (III.4) that

$$\begin{aligned} \forall j \in \llbracket 1, m \rrbracket, \forall k_j \in \llbracket 0, n_j - 1 \rrbracket, \exists k_\Phi \in \llbracket 0, n_\Phi - 1 \rrbracket, \quad \mathcal{L}_j(\hat{x})[k_j] &= \int_{\mathbb{R}} e^{i2\pi \frac{\xi}{f_\Phi}(k_\Phi - \gamma_\Phi)} d\hat{x}(\xi) \\ &= \int_{\mathbb{R}} e^{i2\pi \frac{\xi}{f_\Phi} k_\Phi} d\left(e^{-i2\pi \frac{\xi \gamma_\Phi}{f_\Phi}} \hat{x}(\xi)\right) \\ &= \mathcal{F}_{n_\Phi} \circ \mathcal{M}_{\frac{\gamma_\Phi}{f_\Phi}}(\hat{x})[k]. \end{aligned}$$

Let by $\mathcal{I} \subseteq \llbracket 0, n_\phi - 1 \rrbracket$ the equivalent observation set of the minimal \mathcal{A}_ϕ introduced in Definition III.1 and consider a selection matrix $C_{\mathcal{I}} \in \mathbb{C}^{m \times n_\phi}$ for this set. The above equality ensures the measurement operator admit a factorization of the form

$$\mathcal{L} = C_{\mathcal{I}} \left(\mathcal{F}_{n_\phi, f_\phi} \circ \mathcal{M}_{\frac{\gamma_\phi}{f_\phi}} \right).$$

Finally, $0 \in \mathcal{I}$ by minimality of the grid \mathcal{A}_ϕ , and the selection matrix $C_{\mathcal{I}}$ is an admissible sub-sampling operator in the sense of Definition II.3.

Since any selection matrix $C_{\mathcal{I}} \in \mathbb{C}^{m \times n}$ can be interpreted as a MRSS with m aligned grids taking a single sample ($n_j = 1$ for all $j \in \llbracket 1, m \rrbracket$), the proof of the converse is immediate. ■

APPENDIX C

PROOF OF THEOREM III.5

In both strong and weak condition cases, the proof relies on previous works presented in [1], [3], and is achieved by constructing a polynomial Q_\star satisfying the conditions (II.5). It is been shown in Section III-B that shifting the signal in the time domain leave the dual feasible set invariant, and we will assume without loss of generality that $\gamma_\phi = 0$ so that $\mathcal{L} = C_{\mathcal{I}} \mathcal{F}_n$. Before starting the proof, we introduce the notations

$$\begin{aligned} \Omega_\phi &= \frac{1}{f_\phi} \Xi = \left\{ \frac{\xi}{f_\phi}, \xi \in \Xi \right\} \\ \forall j \in \llbracket 1, p \rrbracket, \quad \Omega_j &= \frac{1}{f_j} \Xi = \left\{ \frac{\xi}{f_j}, \xi \in \Xi \right\} \\ \forall j \in \llbracket 1, p \rrbracket, \quad \tilde{\Omega}_j &= \left\{ \frac{\xi}{f_\phi} + \frac{k}{l_j}, \xi \in \Xi, k \in \llbracket 0, l_j - 1 \rrbracket \right\}. \end{aligned}$$

In the above, Ω_ϕ and Ω_j are the sets of the reduced frequencies of the spectral support Ξ of the signal x for the respective sampling frequencies f_ϕ and f_j , while $\tilde{\Omega}_j$ is the aliased set of Ω_j resulting from a zero-forcing upsampling from the rate f_j to the rate f_ϕ .

We recall from [3] Proposition II.4, using the improved separability conditions taken from [4] Proposition 4.1, that if $\Delta_{\mathbb{T}}(\Omega_j) \geq \frac{2.52}{n_j - 1}$, then one can build a polynomial $P_{j,\star} \in \mathbb{C}^{n_j - 1}[X]$ satisfying the interpolating conditions

$$\begin{cases} P_{j,\star} \left(e^{i2\pi \frac{\xi_r}{f_j}} \right) = \text{sign} \left(e^{i2\pi \frac{a_j}{l_j} \frac{\xi_r}{f_j} \alpha_r} \right), & \forall \frac{\xi_r}{f_j} \in \Omega_j \\ |P_{j,\star} (e^{i2\pi \nu})| < 1, & \forall \nu \in \mathbb{T} \setminus \Omega_j \\ \frac{d^2 |P_{j,\star}|}{d\nu^2} \left(e^{i2\pi \frac{\xi_r}{f_j}} \right) \leq -\eta, & \forall \frac{\xi_r}{f_j} \in \Omega_j, \end{cases} \quad (\text{C.1})$$

provided that $n_j > 2 \times 10^3$, for some $\eta > 0$ ($\eta = 7.865 \cdot 10^{-2}$ in the original proof presented in [4]), and whereby $\{(a_j, l_j)\}_{j \in \llbracket 1, p \rrbracket}$ are the pairs of parameters defined in the statement of Proposition III.2 characterizing the expansion of the array \mathcal{A}_j into the minimal common grid \mathcal{A}_ϕ . If the polynomial $P_{j,\star}$ exists, we further introduce the polynomial $Q_{j,\star} \in \mathbb{C}^{n_\phi - 1}[X]$ defined by

$$\forall z \in \mathbb{C}, \quad Q_{j,\star}(z) = z^{-a_j} P_{j,\star}(z^{l_j}). \quad (\text{C.2})$$

By construction, $Q_{j,\star}$ is a sparse polynomial with monomial support on the subset \mathcal{I} introduced in Proposition III.4. Its coefficients vector $q_{j,\star}$ satisfies the relation $q_{j,\star} = C_{\mathcal{I}}^* c_{j,\star}$ for some $c_{j,\star} \in \mathbb{C}^m$. It is easy to notice that due to the upscaling effect $z \leftarrow z^{l_j}$ in (C.2) the function

$$\begin{aligned} \mathbb{R} &\rightarrow \mathbb{C} \\ \nu &\mapsto |Q_{j,\star}(e^{i2\pi\nu})|, \end{aligned}$$

is $\frac{1}{l_j}$ -periodic. Consequently the polynomial $Q_{j,\star}$ reaches a modulus equal to 1 on every point of $\tilde{\Omega}_j$, with value satisfying

$$\begin{aligned} \forall \nu \in \tilde{\Omega}_j, \quad Q_j(e^{i2\pi\nu}) &= Q_{j,\star}\left(e^{i2\pi\left(\frac{\xi_r}{f_\phi} + \frac{k}{l_j}\right)}\right) \\ &= e^{-i2\pi a_j\left(\frac{\xi_r}{f_\phi} + \frac{k}{l_j}\right)} P_{j,\star}\left(e^{i2\pi\left(\frac{l_j \xi_r}{f_\phi} + k\right)}\right) \\ &= e^{-i2\pi a_j\left(\frac{\xi_r}{f_\phi} + \frac{k}{l_j}\right)} \text{sign}\left(e^{i2\pi \frac{a_j}{l_j} \frac{\xi_r}{f_j} \alpha_r}\right) \\ &= e^{-i2\pi a_j \frac{k}{l_j}} \text{sign}(\alpha_r), \end{aligned}$$

whereby $\frac{\xi_r}{f_\phi} \in \Omega_\phi$ and $k \in \llbracket 0, l_j - 1 \rrbracket$. It comes that the constructed polynomial verifies the interpolation conditions

$$\begin{cases} Q_{j,\star}(e^{i2\pi\nu}) = \text{sign}(\alpha_r), & \forall \nu \in \Omega_\phi \\ Q_{j,\star}(e^{i2\pi\nu}) = e^{-i2\pi a_j \frac{k}{l_j}} \text{sign}(\alpha_r), & \forall \nu \in \tilde{\Omega}_j \\ |Q_{j,\star}(e^{i2\pi\nu})| < 1, & \forall \nu \in \mathbb{T} \setminus \tilde{\Omega}_j \\ \frac{d^2 |Q_{j,\star}|}{d\nu^2}(e^{i2\pi\nu}) \leq -l_j \eta, & \forall \nu \in \tilde{\Omega}_j, \end{cases} \quad (\text{C.3})$$

where the second equality stand for some $\frac{\xi_r}{f_\phi} \in \Omega_\phi$ and $k \in \llbracket 0, l_j - 1 \rrbracket$ such that $\nu = \frac{\xi_r}{f_\phi} + \frac{k}{l_j} \in \tilde{\Omega}_j$.

Under both strong and weak assumptions, we aim to build a sparse polynomial $Q_\star \in \mathbb{C}^{n_\phi-1}[X]$ verifying the conditions (II.5). If the existence of such polynomial is verified II.2 applies and the desired conclusion follows.

Construction under the strong condition: Suppose that $\Delta_{\mathbb{T}}(\Omega_j) \geq \frac{2.52}{n_j-1}$ and $n_j > 2 \times 10^3$, for all $j \in \llbracket 1, p \rrbracket$, as explained above, one can find p polynomials $Q_{j,\star} \in \mathbb{C}^{n_\phi-1}[X]$ satisfying the interpolation properties given in (C.3). Define by $Q_\star \in \mathbb{C}^{n_\phi-1}[X]$ their average

$$\forall z \in \mathbb{C}, \quad Q_\star(z) = \frac{1}{p} \sum_{j=1}^p Q_{j,\star}(z).$$

It is clear, by stability through linear combinations, that Q_\star is still sparse and supported over the subset \mathcal{I} , ensuring the existence of an element $c_\star \in \mathbb{C}^m$ such that $q_\star = C_{\mathcal{I}}^* c_\star$. Moreover, it is immediate to verify that Q_\star satisfies

$$|Q_\star(e^{i2\pi\nu})| = 1 \Leftrightarrow \left(\nu \in \bigcap_{j=1}^p \tilde{\Omega}_j \text{ and } \forall j \in \llbracket 1, p \rrbracket, Q_{j,\star}(e^{i2\pi\nu}) = u(\nu) \right) \quad (\text{C.4})$$

for some value $u(\nu) \in \mathbb{C}$ of modulus 1, $|u(\nu)| = 1$. Let us denote by $\Gamma \subset \mathbb{T}$ the set of frequencies satisfying (C.4). From (II.5) and (C.3), Q_\star is a dual certificate if and only if $\Gamma = \Omega_\phi$. One has $\Omega_\phi \subseteq \Gamma$, thus it remains to

prove $\Gamma \subseteq \Omega_\phi$ to finish the certificate construction under the strong condition. Using the definition of $\tilde{\Omega}_j$ and the interpolation properties (C.3), we have that $\nu \in \Gamma$ is equivalent to

$$\nu \in \bigcap_{j=1}^p \tilde{\Omega}_j \iff \forall (j, j') \in \llbracket 1, p \rrbracket^2, \exists (r, r') \in \llbracket 1, s \rrbracket^2, \exists k_j \in \llbracket 0, l_j - 1 \rrbracket, \exists k_{j'} \in \llbracket 0, l_{j'} - 1 \rrbracket, \\ e^{-i2\pi a_j \frac{k_j}{l_j} \text{sign}(\alpha_r)} = e^{-i2\pi a_{j'} \frac{k_{j'}}{l_{j'}} \text{sign}(\alpha_{r'})},$$

leading to

$$\nu \in \bigcap_{j=1}^m \tilde{\Omega}_j \iff \forall (j, j') \in \llbracket 1, p \rrbracket^2, \exists (r, r') \in \llbracket 1, s \rrbracket^2, \exists k_j \in \llbracket 0, l_j - 1 \rrbracket, \exists k_{j'} \in \llbracket 0, l_{j'} - 1 \rrbracket, \exists b \in \mathbb{Z}, \\ a_j \frac{k_j}{l_j} + \frac{\arg(\alpha_r)}{2\pi} = a_{j'} \frac{k_{j'}}{l_{j'}} + \frac{\arg(\alpha_{r'})}{2\pi} + b. \quad (\text{C.5})$$

The equality in the RHS of (C.5) may occur for all pairs $(j, j') \in \llbracket 1, p \rrbracket^2$ if and only if $r = r'$, and the above reduces to

$$\nu \in \bigcap_{j=1}^m \tilde{\Omega}_j \iff \forall (j, j') \in \llbracket 1, p \rrbracket^2, \exists k \in \llbracket 0, l_j - 1 \rrbracket, \exists k' \in \llbracket 0, l_{j'} - 1 \rrbracket, \exists b \in \mathbb{Z}, \quad \frac{a_j k_j}{l_j} = \frac{a_{j'} k_{j'}}{l_{j'}} + b,$$

which holds if and only if

$$\forall (j, j') \in \llbracket 1, p \rrbracket^2, \exists k \in \llbracket 0, l_j - 1 \rrbracket, l_j \mid a_j l_{j'} k_j.$$

Recalling from the minimality condition of the common grid \mathcal{A}_ϕ detailed in Proposition III.2 that $\gcd(\{a_j\}_{j \in \llbracket 1, p \rrbracket} \cup \{l_j\}_{j \in \llbracket 1, p \rrbracket}) = 1$, one derives by application of the Gauss theorem

$$\exists j \in \llbracket 1, p \rrbracket, \quad l_j \mid k_j.$$

Since $k_j \in \llbracket 0, l_j - 1 \rrbracket$, one has $k_j = 0$. We deduce that there must exists $r \in \llbracket 1, s \rrbracket$ such that $\nu = \frac{\xi_r}{f_\phi} + \frac{0}{l_j}$ and finally $\nu \in \Omega_\phi$. Consequently, $\Gamma \subseteq \Omega_\phi$, and finally $\Gamma = \Omega_\phi$, which concludes the proof for the strong condition. ■

Construction under the weak condition: Suppose that $\Delta_{\mathbb{T}}(\Omega_j) \geq \frac{2.52}{n_j - 1}$ and $n_j > 2 \times 10^3$ for some $j \in \llbracket 1, p \rrbracket$, and define the polynomial $Q_{j,*} \in \mathbb{C}^{n_\phi - 1}[X]$ as in Equation (C.2). Moreover, we define by $\mathcal{H}_j(\mathbb{A}, \Omega_\phi)$ the affine subspace of elements $c \in \mathbb{C}^m$ such that $q = C_{\mathcal{I}}^* c$ induces a sparse polynomial $Q \in \mathbb{C}^{n_\phi - 1}[X]$ supported by monomials taken over the subset \mathcal{I} and satisfying the interpolation conditions

$$\begin{cases} Q(e^{i2\pi\nu}) = \text{sign}(\alpha_r), & \forall \nu \in \Omega_\phi \\ Q'(e^{i2\pi\nu}) = 0, & \forall \nu \in \Omega_\phi \\ Q(e^{i2\pi\nu}) = 0, & \forall \nu \in \tilde{\Omega}_j \setminus \Omega_\phi. \end{cases}$$

The subspace $\mathcal{H}_j(\mathbb{A}, \xi)$ can be parametrized by the linear equality

$$\mathcal{H}_j(\mathbb{A}, \xi) = \{c \in \mathbb{C}^m, V_j(\mathbb{A}, \Omega_\phi) C_{\mathcal{I}}^* c = w\},$$

whereby $w = [\text{sign}(\alpha_1), \dots, \text{sign}(\alpha_s)]^\top \in \mathbb{C}^s$, and for some matrix $V_j(\mathbb{A}, \Omega_\Phi) \in \mathbb{C}^{(l_j+1)s \times n_\Phi}$ defining the interpolation conditions. Interpolation theory guarantees that $V_j(\mathbb{A}, \Omega_\Phi)$ is full rank, and therefore the subspace $\mathcal{H}_j(\mathbb{A}, \Omega_\Phi)$ is non-trivial with dimension $m - (l_j + 1)s$, provided that $m \geq (l_j + 1)s$. We fix an element $t \in \mathcal{H}_j(\mathbb{A}, \Omega_\Phi)$, and denote by $R \in \mathbb{C}^{n_\Phi-1}[X]$ the polynomial having for coefficients vector $r = C_{\mathcal{T}}^* t$. In the rest of this proof, we seek to build a dual certificate $Q_\star \in \mathbb{C}^{n_\Phi-1}[X]$ under the form of a convex combination between R and $Q_{j,\star}$

$$Q_\star = \beta R + (1 - \beta) Q_{j,\star}, \quad \beta \in [0, 1].$$

First of all, by construction, R and $Q_{j,\star}$ both interpolate the frequencies of Ω_Φ with values $w_r = \text{sign}(a_r)$, and one has

$$\forall \nu \in \Omega_\Phi, \quad Q_\star(e^{i2\pi\nu}) = w_r. \quad (\text{C.6})$$

Consequently, it remains to derive sufficient conditions on β for the optimality condition $|Q_\star(e^{i2\pi\nu})| < 1$ to hold everywhere else on $\mathbb{T} \setminus \Omega_\Phi$ to ensure that Q_\star is a dual certificate. To do so, we partition the set \mathbb{T} into three non-intersecting sets $\mathbb{T} = \Gamma_{\text{near}} \cup \Gamma_{\text{alias}} \cup \Gamma_{\text{far}}$, where Γ_{near} is a union of s open ball of small radii $0 < \varepsilon_{\text{near}}$ centered around the frequencies in Ω_Φ , Γ_{alias} is an open set containing the elements of $\tilde{\Omega}_j \setminus \Omega_\Phi$. The set Γ_{far} is defined by the complementary of the two previous in \mathbb{T} . The conditions on β for Q_\star to be bounded away from 1 in modulus are derived independently on each of those sets.

We start the analysis on Γ_{near} . For any complex polynomial Q , we respectively denote by $Q_{\Re}(\nu) = \Re(Q(e^{i2\pi\nu}))$ and $Q_{\Im}(\nu) = \Im(Q(e^{i2\pi\nu}))$ for all $\nu \in \mathbb{T}$, its real and imaginary part around the unit circle. Moreover, we recall that

$$\frac{d^2 |Q|}{d\nu^2}(\nu) = -\frac{(Q_{\Re}(\nu) Q'_{\Re}(\nu) + Q_{\Im}(\nu) Q'_{\Im}(\nu))^2}{|Q(\nu)|^3} + \frac{|Q'(\nu)|^2 + Q_{\Re}(\nu) Q''_{\Re}(\nu) + Q_{\Im}(\nu) Q''_{\Im}(\nu)}{|Q(\nu)|}, \quad (\text{C.7})$$

for all $\nu \in \mathbb{T}$. By construction, the derivative of R and $Q_{j,\star}$ cancels on Ω_Φ and by linearity

$$\forall \nu \in \Omega_\Phi, \quad Q'_\star(e^{i2\pi\nu}) = 0. \quad (\text{C.8})$$

Injecting Equations (C.6) and (C.8) into (C.7) leads to

$$\forall \nu \in \Omega_\Phi, \quad \frac{d^2 |Q_\star|}{d\nu^2}(\nu) = \cos(w_r) Q''_{\star\Re}(\nu) + \sin(w_r) Q''_{\star\Im}(\nu).$$

Thus, the operator $\frac{d^2 |\cdot|}{d\nu^2}$ acts linearly on the polynomial Q_\star at the points in Ω_Φ , and one has

$$\begin{aligned} \forall \nu \in \Omega_\Phi, \quad \frac{d^2 |Q_\star|}{d\nu^2}(\nu) &= \beta \frac{d^2 |R|}{d\nu^2}(\nu) + (1 - \beta) \frac{d^2 |Q_{j,\star}|}{d\nu^2}(\nu) \\ &\leq \beta \frac{d^2 |R|}{d\nu^2}(\nu) - (1 - \beta) l_j \eta, \end{aligned}$$

using the interpolation properties of Equation (C.3). The inequalities

$$\forall \nu \in \Omega_\Phi, \quad \frac{d^2 |Q_\star|}{d\nu^2}(\nu) < 0$$

can be jointly satisfied, for a choice of β

$$\beta < \frac{l_j \eta}{\mathcal{M}''_\Phi(R) + l_j \eta}, \quad (\text{C.9})$$

where

$$\mathcal{M}_\phi''(R) = \max_{\nu \in \Omega_\phi} \frac{d^2 |R|}{d\nu^2}(\nu).$$

Under Condition (C.9), $|Q_\star| - 1$ has s non-nodal roots on Ω_ϕ , and by continuity of Q_\star there must exist a radius $0 < \varepsilon_{\text{near}}$ such that

$$\forall \nu \in \Gamma_{\text{near}} \setminus \Omega_\phi, \quad |Q_\star(e^{i2\pi\nu})| < 1,$$

holds where $\Gamma_{\text{near}} = \bigcup_{r=1}^s \mathcal{B}\left(\frac{\xi_r}{f_\phi}, \varepsilon_{\text{near}}\right)$, where $\mathcal{B}(\nu, \varepsilon)$ denotes the open ball of \mathbb{T} of center ν and radius ε for the torus distance.

We continue the proof by bounding $|Q_\star|$ away from 1 on the set Γ_{alias} . Fix any $0 < \delta < 1$ and let $\Gamma_{\text{alias}} = \{\nu, |R(e^{i2\pi\nu})| < \delta\}$. By continuity of R , Γ_{alias} is an open set verifying $(\tilde{\Omega}_j \setminus \Omega_\phi) \subset \Gamma_{\text{alias}}$, moreover one can impose $\Gamma_{\text{alias}} \cap \Gamma_{\text{near}} = \emptyset$ for a small enough δ . The value of $|Q_\star|$ over Γ_{alias} can be bounded by

$$\begin{aligned} \forall \nu \in \Gamma_{\text{alias}}, \quad |Q_\star(e^{i2\pi\nu})| &\leq \beta |R(e^{i2\pi\nu})| + (1 - \beta) |Q_{j,\star}(e^{i2\pi\nu})| \\ &< \beta\delta + (1 - \beta). \end{aligned}$$

Consequently, $|Q|$ is smaller than 1 on Γ_{alias} as long as $\beta > 0$.

It remains to prove that $|Q|$ can also be bounded by 1 in the rest of the torus $\Gamma_{\text{far}} = \overline{\mathbb{T} \setminus (\Gamma_{\text{true}} \cup \Gamma_{\text{alias}})}$. Let by $\mathcal{M}_{\text{far}}(R)$ and $\mathcal{M}_{\text{far}}(Q_{j,\star})$ be the respective suprema of R and $Q_{j,\star}$ over Γ_{far} . Γ_{far} is a closed set, and thus compact. It comes that the suprema of R and Q are reached in some points inside Γ_{far} . Moreover introducing the suprema of $Q_{j,\star}$ over this set

$$\mathcal{M}_{\text{far}}(Q_{j,\star}) = \sup_{\nu \in \Gamma_{\text{far}}} \{|Q_{j,\star}(e^{i2\pi\nu})|\} < 1,$$

since $\tilde{\Omega}_j \not\subset \Gamma_{\text{far}}$, leads to

$$\begin{aligned} \forall \nu \in \Gamma_{\text{far}}, \quad |Q_\star(e^{i2\pi\nu})| &\leq \beta |R(e^{i2\pi\nu})| + (1 - \beta) |Q_{j,\star}(e^{i2\pi\nu})| \\ &< \beta \mathcal{M}_{\text{far}}(R) + (1 - \beta) \mathcal{M}_{\text{far}}(Q_{j,\star}) \end{aligned}$$

for all $\nu \in \Gamma_{\text{far}}$, and thus $|Q_\star(e^{i2\pi\nu})| < 1$ can be achieved everywhere on Γ_{far} provided a choice of β verifying

$$\beta < \frac{1 - \mathcal{M}_{\text{far}}(Q_{j,\star})}{\mathcal{M}_{\text{far}}(R) - \mathcal{M}_{\text{far}}(Q_{j,\star})}.$$

We conclude that for any coefficient β satisfying

$$0 < \beta < \min \left\{ \frac{l_j \eta}{\mathcal{M}_\phi''(R) + l_j \eta}, \frac{1 - \mathcal{M}_{\text{far}}(Q_{j,\star})}{\mathcal{M}_{\text{far}}(R) - \mathcal{M}_{\text{far}}(Q_{j,\star})} \right\},$$

the polynomial Q_\star meet the conditions (II.5) and thus qualifies as a dual certificate. \blacksquare

APPENDIX D

PROOF OF PROPOSITION III.2

A. Existence of a common grid

Suppose that \mathcal{A}_+ is a common supporting grid for the set of arrays \mathbb{A} . Relation (III.4) ensures

$$\forall j \in \llbracket 1, p \rrbracket, \quad \forall k \in \llbracket 0, n_j - 1 \rrbracket, \quad \exists q_j[k] \in \llbracket 0, n_+ - 1 \rrbracket \quad \text{s.t.} \quad \frac{1}{f_j}(k - \gamma_j) = \frac{1}{f_+}(q_j[k] - \gamma_+), \quad (\text{D.1})$$

whereby each integer $q_j[k]$ represents the position of the k^{th} samples of the j^{th} grid in the common grid. By subtracting two instances of (D.1) applied to the grid j and for the samples of order k and $k+1$ one gets

$$\forall j \in \llbracket 1, p \rrbracket, \forall k \in \llbracket 0, n_j - 1 \rrbracket, \quad \frac{f_+}{f_j} = q_j[k+1] - q_j[k] \triangleq l_j,$$

where $\{l_j\}_{j \in \llbracket 1, p \rrbracket}$ are positive integers since q_j is an increasing sequence for all $j \in \llbracket 1, p \rrbracket$. It comes that $\{q_j\}_{j \in \llbracket 1, p \rrbracket}$ are p arithmetic progressions with respective increment l_j

$$\forall j \in \llbracket 1, p \rrbracket, \forall k \in \llbracket 0, n_j - 1 \rrbracket, \quad q_j[k] = q_j[0] + l_j k.$$

Reporting those results in Equation (D.1) leads to

$$\forall j \in \llbracket 1, p \rrbracket, \quad \gamma_+ = q_j[0] + l_j \gamma_j.$$

Letting $a_j = -q_j[0]$ for all $j \in \llbracket 1, p \rrbracket$ proofs the necessity part.

On the other hand, suppose now the existence of positive integers $\{l_j\} \in \mathbb{N}^p$ and integers $\{a_j\} \in \mathbb{Z}^p$ such that the relations

$$\begin{cases} f_+ = l_j f_j, & \forall j \in \llbracket 1, p \rrbracket \\ \gamma_+ = l_j \gamma_j - a_j, & \forall j \in \llbracket 1, p \rrbracket, \end{cases} \quad (\text{D.2})$$

hold for some $f_+ \in \mathbb{R}^+$ and $\gamma_+ \in \mathbb{R}$. It comes

$$\begin{aligned} \forall j \in \llbracket 1, p \rrbracket, \forall k \in \llbracket 0, n_j - 1 \rrbracket, \quad \frac{1}{f_j} (k - \gamma_j) &= \frac{1}{f_j} (k - l_j a_j - l_j \gamma_+) \\ &= \frac{1}{f_+} (l_j k - a_j - \gamma_+). \end{aligned} \quad (\text{D.3})$$

Defining the quantities

$$\begin{cases} q_j[k] = l_j k - a_j, & \forall j \in \llbracket 1, p \rrbracket \\ n_+ \geq \max_{j \in \llbracket 1, p \rrbracket} \{q_j[n_j - 1]\}, \end{cases} \quad (\text{D.4})$$

ensures that the grid $\mathcal{A}_+ = (f_+, \gamma_+, n_+)$ supports the system defined by \mathbb{A} . This achieves the sufficiency part, and thus the characterization of the existence of a common grid.

B. Conditions for minimality

Suppose that \mathbb{A} admits a common grid, it is clear that exactly one element of $\mathcal{C}(\mathbb{A})$ reaches the minimal order n_Φ . Denote by $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ this element. Moreover, denote by $\{l_j\} \in \mathbb{N}^p$ and $\{a_j\} \in \mathbb{Z}^p$ the elements characterizing the grid expansion of \mathbb{A} onto \mathcal{A}_Φ defined in (D.2), and let $\delta = \gcd(\{a_j\}_{j \in \llbracket 1, p \rrbracket} \cup \{l_j\}_{j \in \llbracket 1, p \rrbracket})$. By (D.3), one has

$$\forall j \in \llbracket 1, p \rrbracket, \forall k \in \llbracket 0, n_j - 1 \rrbracket, \quad \frac{1}{f_j} (k - \gamma_j) = \frac{\delta}{f_\Phi} \left(\frac{l_j}{\delta} k - \frac{a_j}{\delta} - \frac{\gamma_\Phi}{\delta} \right),$$

Thus the grid $\mathcal{A}_\Phi = \left(\frac{f_\Phi}{\delta}, \frac{\gamma_\Phi}{\delta}, \left\lceil \frac{n_\Phi}{\delta} \right\rceil \right)$ supports \mathbb{A} and belongs to $\mathcal{C}(\mathbb{A})$. My minimality of \mathcal{A}_Φ one has $\left\lceil \frac{n_\Phi}{\delta} \right\rceil \geq n_\Phi$ and we conclude that $\delta = 1$. Moreover, the minimality implies that the first and the last samples of the grid \mathcal{A}_Φ must be

acquired by an element of \mathbb{A} , otherwise the shorter grids $\mathcal{A}_\phi = (f_\phi, \gamma_\phi - 1, n_\phi - 1)$, or $\mathcal{A}_\phi = (f_\phi, \gamma_\phi + 1, n_\phi - 1)$ would also support \mathbb{A} . Using (D.4)

$$\begin{cases} \forall j \in \llbracket 1, p \rrbracket, & \gamma_\phi = l_j \gamma_j - a_j \\ \exists j \in \llbracket 1, p \rrbracket, & a_j = 0 \\ \forall j \in \llbracket 1, p \rrbracket, & a_j \leq 0, \end{cases}$$

which implies $\gamma_\phi = \max_{j \in \llbracket 1, p \rrbracket} \{l_j \gamma_j\}$, ensuring that the conditions describing the minimal grid stated in Proposition III.2 are necessary.

For the sufficiency, consider the grid $\mathcal{A}_\phi = (f_\phi, \gamma_\phi, n_\phi)$ of $\mathcal{C}(\mathbb{A})$ where $\gamma_\phi = \max_{j \in \llbracket 1, p \rrbracket} \{l_j \gamma_j\}$ and with expansion parameters $\{l_j\} \in \mathbb{N}^p$ and $\{a_j\} \in \mathbb{Z}^p$ satisfying $\gcd(\{a_j\} \cup \{l_j\}, j \in \llbracket 1, p \rrbracket) = 1$. Let $\mathcal{A}' = (f', \gamma', n') \in \mathcal{C}(\mathbb{A})$ be any other grid and let by δ' its corresponding greatest common divisor. δ' divides every integer linear combination of $\{a_j\} \cup \{l_j\}$, and in particular every elements of the set $\{l_j k_j - a_j : j \in \llbracket 1, p \rrbracket, k_j \in \llbracket 0, n_j - 1 \rrbracket\}$. Therefore (f', γ') is identifiable to $(\delta' f_\phi, \delta' \gamma_\phi - b)$ for some $b \in \mathbb{Z}$. Moreover since γ_ϕ is maximum, the grid \mathcal{A}_ϕ samples an element of \mathbb{A} at index 0, and thus $\mathcal{A}' \in \mathcal{C}(\mathbb{A})$ if and only if $b \geq 0$. Finally it comes from (D.4) that n' must satisfy

$$\begin{aligned} n' &\geq \max_{j \in \llbracket 1, p \rrbracket} \{q'_j [n_j - 1]\} \\ &\geq \max_{j \in \llbracket 1, p \rrbracket} \{\delta' l_j (n_j - 1) - \delta' l_j a_j + b\} \\ &\geq \max_{j \in \llbracket 1, p \rrbracket} \{l_j (n_j - 1) - a_j\} \\ &\geq n_\phi, \end{aligned}$$

demonstrating the sufficiency part, and concluding the proof of Proposition III.2.

REFERENCES

- [1] E. J. Candès and C. Fernandez-Granda, “Towards a mathematical theory of super-resolution,” *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.
- [2] —, “Super-resolution from noisy data,” *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.
- [3] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, “Compressed sensing off the grid,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.
- [4] C. Fernandez-Granda, “Super-resolution of point sources via convex programming,” *arXiv:1507.07034*, 2015.
- [5] P. Stoica and A. Nehorai, “MUSIC, maximum likelihood, and Cramer-Rao bound,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.
- [6] R. Roy and T. Kailath, “ESPRIT - Estimation of signal parameters via rotational invariance techniques,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [7] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse sampling of signal innovations,” *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 31–40, Mar. 2008.
- [8] G. Tang, B. N. Bhaskar, and B. Recht, “Near minimax line spectral estimation,” *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, Jan. 2015.
- [9] B. N. Bhaskar, G. Tang, and B. Recht, “Atomic norm denoising with applications to line spectral estimation,” *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.

- [10] V. Duval and G. Peyré, “Exact support recovery for sparse spikes deconvolution,” *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.
- [11] T. Bendory, S. Dekel, and A. Feuer, “Robust recovery of stream of pulses using convex optimization,” *Journal of Mathematical Analysis and Applications*, vol. 442, no. 2, pp. 511 – 536, 2016.
- [12] G. Tang, “Resolution limits for atomic decompositions via Markov-Bernstein type inequalities,” *2015 International Conference on Sampling Theory and Applications, SampTA 2015*, pp. 548–552, May 2015.
- [13] G. Schiebinger, E. Robeva, and B. Recht, “Superresolution without separation,” *arXiv:1506.03144*, 2015.
- [14] G. Tang, B. N. Bhaskar, and B. Recht, “Sparse recovery over continuous dictionaries-just discretize,” in *Conference Record - Asilomar Conference on Signals, Systems and Computers*, no. 4. IEEE, Nov. 2013, pp. 1043–1047.
- [15] N. Boyd, G. Schiebinger, and B. Recht, “The alternating descent conditional gradient method for sparse inverse problems,” *arXiv:1507.01562*, 2015.
- [16] Y. Chi and Y. Chen, “Compressive two-dimensional harmonic retrieval via atomic norm minimization,” *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1030–1042, Feb. 2015.
- [17] Z. Yang, L. Xie, and P. Stoica, “Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution,” *arXiv:1505.02510*, 2015.
- [18] Y. Li and Y. Chi, “Off-the-grid line spectrum denoising and estimation with multiple measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, Mar. 2016.
- [19] Z. Yang and L. Xie, “Exact joint sparse frequency recovery via optimization methods,” *arXiv:1405.6585*, 2014.
- [20] D. Yang, G. Tang, and M. B. Wakin, “Super-resolution of complex exponentials from modulations with unknown waveforms,” *arXiv:1601.03712*, 2016.
- [21] J.-F. Cai, X. Qu, W. Xu, and G.-B. Ye, “Robust recovery of complex exponential signals from random gaussian projections via low rank hankel matrix reconstruction,” *Applied and Computational Harmonic Analysis*, vol. 41, no. 2, pp. 470–490, 2016.
- [22] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, “Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach,” *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4649–4662, Sep. 2016.
- [23] R. Heckel, V. I. Morgenshtern, and M. Soltanolkotabi, “Super-resolution radar,” *Information and Inference*, vol. 5, no. 1, pp. 22–75, Mar. 2016.
- [24] P. P. Vaidyanathan and P. Pal, “Sparse sensing with co-prime samplers and arrays,” *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, Feb. 2011.
- [25] J. F. Sturm, “Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones,” *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 625–653, Aug. 1999.
- [26] K. C. Toh, M. J. Todd, and R. H. Tütüncü, “SDPT3 A Matlab software package for semidefinite programming, version 1.3,” *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 545–581, Aug. 2001.
- [27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [28] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*. Springer, 2010.

Maxime Ferreira Da Costa

Wei Dai